أفضل الكتب مبيعًا لدى نيويورك تايمز

الإحصاء عاريًا

أن نُنزع عن البيانات مهابتها



مؤلف كتاب «الاقتصاد عاريًا»

تشارلز ويلان

«كتاب رائع ماتع ... هو معلم الرياضيات الأفضل الذي لم تدرس على يديه من قبل».

- صحيفة سان فرانسيسكو كرونيكل

الإحصاء عاريًا

الإحصاء عاريًا
تشارلز ويلان

Naked Statistics
Charles Wheelan

الإحصاء عاريًا
Naked Statistics
تشارلز ويلان
Charles Wheelan

الطبعة الأولى: ٢٠٢١م
رقم إيداع: ١٦٠٨٣/ ٢٠٢٠
تدمك: ٩٧٨٩٧٧٨٥٧٦٧١٩
٢٨٠ص، ١٦×٢٣سم

جمهورية مصر العربية
٦٦ مساكن الرماية، الدور الثالث - شقة ١٠، الهرم، الجيزة

+201099596575
bookmania2017@gmail.com
بوك مانيا - Bookmania
book_mania2017

تصميم وتنفيذ
شركة خطوة

الإحصاء عاريًا

أن تُنزع عن البيانات مهابتها

تأليف

تشارلز ويلان

ترجمة

شهاب ياسين



بوك مانيا

المحتويات

مزيد من المديح في كتاب

الإحصاء عاريًا

لكاتبه تشارلز ويلان

«كتاب جديد حاذق ماتع ... إن ويلان معلم رياضيات ذو براعة لم تشهدها من قبل؛ لذا فهو لن يغرقك في التفاصيل الفنية ... فالكتاب مفعم بالدروس العملية من نوعية: كيفية الحكم على صحة استطلاعات الرأي، والأسباب التي ستجعلك تعزف عن ابتياع تذاكر اليانصيب للأبد، وكيفية رصد إشارات الخطر التي تظهر في البيانات العامة».

صحيفة سان فرانسيسكو كرونيكل

«لطالما كانت كتب الإحصاء، وقضاء وقت ممتع عبارتين لا تراهما متصاحبتين عادة، إلى أن نجح تشارلي ويلان في الجمع بينهما في هذا الكتاب. ففيه يبين لك كيف تساعدك الأفكار الإحصائية على فهم الكثير من الأمور المتعلقة بالحياة اليومية».

أوستان جولسبي، أستاذ الاقتصاد بجامعة شيكاجو
والرئيس السابق لمجلس المستشارين الاقتصاديين

«كتاب ممهد للإحصاء، مشوق مكتوب بحرفية، جاء ممتعًا على غير ما هو متوقع يصعب على المرء تخيل وجود كتاب تمهيدي مماثل يقدم لذلك المجال ذي السمعة التي لا يستحقها، من كونه مجالًا يتعصى على الفهم».

مجلة نيو ريبابليك

«كتاب ممتع جاذب للاهتمام يبين أهمية الإحصاء كأداة مهمة، بها يستطيع الناس أن يفهموا العالم المعاصر».

جيكوب جيه جولدستين
أحد مقدمي برنامج بلانيت ماني على إذاعة إن بي آر

«لقد قام تشارلز ويلان، الأستاذ بكلية دارتموث (والمراسل السابق لجريدة الإيكونوميست بشيكاجو) بأمر فريد من نوعه؛ إذ جعل الإحصاء أمرًا ممتعًا جاذبًا للاهتمام. فكتابه هذا يجرد الموضوع من تعقيداته ليكشف عما تحتها مما يخلب اللب وهو كتاب يستمد نجاحه من العلامة التجارية الفكاهية التي صكَّها ويلان (واستخدمها في كتابه السابق الاقتصاد عاريًا) كيلا يجعل القارئ يفقد اهتمامه بالكتاب ويرتاح لقراءته. ولقد تمخض عن جهود ويلان نتيجة رائعة … فهذا الكتاب هو نظرة كاشفة لأساسيات الإحصاء مجردة».

مجلة الإيكونوميست

«هل أنت ممن يهابون علم الإحصاء؟ فلتتوقف عن ذلك. فكتاب تشارلز ويلان هذا يبين المنطقية التي تكمن وراء تلك المفاهيم الإحصائية المتعددة التي نستخدمها بطريقة سهلة طيعة».

راجورام راجان مؤلف كتاب
Fault Lines: How Hidden Fractures Still
Threaten the World Economy

«مستخدمًا الفكاهة والأسلوب الحواري الجذاب، يقود ويلان القارئ خلال أساسيات المفاهيم الإحصائية وتطبيقاتها عن طريق أمثلة من العالم الواقعي توضح كيف تعمل الإحصائيات ولماذا هي مهمة. إنه كتاب ممتاز في مجمله».

مجلة ساينس نيوز

«لَكَمْ تمنيت أن أحظى بهذا الكتاب عام ١٩٩١ لأنني في ذلك العام درست الإحصاء لأول مرة في أولى فصولي الدراسية في كلية الدراسات العليا … إن ويلان أستاذ بارع في تبيان المبادئ الرئيسية للإحصائيات ومناهجها بطريقة سهلة يألفها القارئ. فهو ولا شك معلم بارع تتبدَّى مهاراته جلية في هذا الكتاب».

مجلة إنسايد هاير إيد الإلكترونية

«هل أردت يومًا أن تختزل مجمل أداء أحد لاعبي مركز خط الوسط في كرة القدم الأمريكية في رقم مفرد، أو تساءلت إذا ما كانت نتيجة إحدى مباريات البولينج تعتبر إحصائية وصفية، أو كنت تأمل في إيجاد تحليل رياضي لمسألة اختيارات لعبة مونتي هول؟ إذن فهذا هو الكتاب الذي تبحث عنه».

مجلة نيتشر

«لقد أتى كتاب ويلان السهل الممتع هذا في وقته تمامًا».

مجلة ساينس

«يقدم هذا الكتاب يَدَ عَوْنٍ ونظرة فكاهية لكل من جابهوا الإحصاء فخامرهم الارتباك أو شعروا بالخداع والضياع ... إنه كتاب ماتع منير».

مجلة بابليشرز ويكلي

«كتاب آسر مفيد مثير للاهتمام».

مجلة نيو ساينتست

«لقد شن ويلان هجومًا ناجحًا على كل تلك الرهبة التي تحيط الإحصاء، فعند قراءتك لهذا الكتاب ستستمتع وتتنور بذات القدر».

مجلة تايمز هاير إيدوكيشن

«دليل ممتع مفعم بالمعلومات عن ذلك الموضوع الباعث على الرهبة دومًا».

مجلة كيركس ريفيوز

مقدمة

لماذا كرهت حساب التفاضل والتكامل وأحببت علم الإحصاء

لطالما جمعتني بالرياضيات علاقة يخيِّم عليها جوٌّ من عدم الارتياح، فإنني لا أحب الأرقام لمجرد كونها أرقامًا. ولا تثير تلك المعادلات المنمقة، التي ليس لها من تطبيقات في العالم الواقعي، اهتمامي. بل إنني كرهت مادة التفاضل والتكامل التي دَرَستُها في المرحلة الثانوية؛ تحديدًا، لأن أحدًا لم يكلف نفسه عناء أن يبيِّن لي ما حاجتي بتعلمها. ما تلك المنطقة التي تقع أسفل القطع المكافئ؟ ومن عساه يبالي بها؟

لكنني، في الواقع، قد مررت بإحدى أعظم لحظات حياتي خلال سَنَتِي النهائية في المدرسة الثانوية مع نهاية الفصل الدراسي الأول من برنامج التفاضل والتكامل المتقدم. كنت أدرُس للاختبار النهائي مقرًّا بأنني لست مستعدًّا لذلك بما يكفي. (كنت قد قُبلتُ في الكلية التي وضعتُها على رأس اختياراتي الدراسية قبل ذلك بأسابيع قليلة، وهو الأمر الذي استنزف كل ما كان لديَّ من دافع هزيل لدراسة تلك المادة). وبينما كنت أطالع أسئلة الاختبار النهائي لم يبدُ أيٌّ منها مألوفًا لديَّ. لا أعني هنا أنني لم أكن قادرًا على الإجابة عنها، وإنما أعني أنني لم أكن حتى أعرف ما المطلوب منها. لم تكن تلك المرة الأولى التي أخضع فيها لاختبار وأنا غير مستعد له، لكن، ودعوني هنا أعيد صياغة عبارة دونالد رمسفيلد*، فأقول إنني كنت دائمًا أعرف ما لا أعرف. بدا لي هذا الاختبار أكثر احتواءً على الأحرف اليونانية عما هو معتاد. فأخذت أقلِّب صفحات الاختبار لبرهة ثم ما لبثت أن استسلمت بشكل أو بآخر. فنهضت ومشيت قاصدًا مقدمة قاعة الدراسة حيث كانت معلمة التفاضل والتكامل التي درَّستني المادة، والتي سنسميها السيدة كارول سميث وكانت تراقب عملية الاختبار. قلت لها: «سيدة سميث، إنني لا أتبين كثيرًا مما ورد في هذا الامتحان».

هنا يجب أن أشير إلى أنني لم أكن أروق للسيدة سميث ولا كانت هي تروقني.

---

* وزير الدفاع الأمريكي الأسبق.

والآن أستطيع أن أعترف بأنني أحيانًا ما كنت أستغل سلطتي المحدودة بصفتي رئيسًا لاتحاد الطلبة كي أخطط لجميع اللقاءات الطلابية بشكل يؤدي إلى إلغاء صفوف حساب التفاضل والتكامل التي تدرسها السيدة سميث. والآن أعترف أنني ورفاقي كنا نرسل لها زهورًا في قاعة الدراسة بطاقتُها ممهورة بتوقيع معجب سري كي نشرع في الضحك مهللين من مؤخرة القاعة بينما تتلفت هي حولها محرجة. والآن أعترف أنني توقفت عن أداء جميع واجباتي الدراسية ما إن تلقيت خبر قبولي في الكلية.

لذلك فإنني عندما ذهبت إليها في منتصف الاختبار وقلت لها إن ما في الاختبار لا يبدو مألوفًا لي، وجدتُها، في الواقع، غير متعاطفة معي. بل إنها صاحت بصوت عالٍ توجِّه الكلام لي ظاهريًّا، ولكن تواجه الجالسين على مقاعد الصفوف الأولى كي تُسمعهم ما تقول: «لو كنتَ قد ذاكرت دروسك يا تشارلز، لما بدت لك الأسئلة غريبة لهذا الحد». كان منطقها سليمًا لا يمكن الجدال معه.

هكذا قفلتُ عائدًا لمقعدي. لكن بعد دقائق قليلة وجدت برايان آربيتر، ذلك الطالب الذي يتفوق عليَّ كثيرًا في حساب التفاضل والتكامل، يمشي متوجهًا إلى حيث السيدة سميث ثم يهمس لها بكلمات قليلة فتجيبه هي هامسة كذلك. ثم حدث أمر في غاية العجب إذ قامت السيدة سميث وأعلنت بصوت عالٍ قائلة: «أيها التلاميذ، أعيروني انتباهكم، يبدو أنني أعطيتكم اختبار الفصل الدراسي الثاني عن طريق الخطأ». كان وقت الاختبار قد استُهلك منه قدر كبير مما حتم إلغاءه وإعادته في وقت آخر.

لا أستطيع أن أصف لكم السعادة الكبيرة التي غمرتني في تلك اللحظة. إنني قد تزوجت بامرأة رائعة وأنجبت ثلاثة أبناء أصحاء، ونشرت العديد من الكتب بِاسمي، وزرت أماكن مثل تاج محل وأنجكور وات؛ لكن تلك اللحظة التي نالت فيها معلمتي ذلك القصاص العادل هي واحدة من أعظم خمس لحظات مررت بها في حياتي. (وهذه التجربة الحياتية الرائعة لم يفسدها كوني كدت أن أرسب في الاختبار بعد أن تمت إعادته في وقت لاحق).

إن واقعة التفاضل والتكامل تلك ستفصح لك عن الكثير، لا كل شيء، مما تحتاج لمعرفته عن علاقتي بالرياضيات. الغريب أنني أحببت مادة الفيزياء في المدرسة الثانوية رغم أنها تقوم على قدر كبير من ذات حسابات التفاضل والتكامل التي زهدتها في صف السيدة سميث. لماذا؟ لأن للفيزياء غاية واضحة. إنني لا أزال أذكر، على وجه التحديد، كيف كان معلم الفيزياء في المدرسة الثانوية يشرح لنا كيف يمكننا أن نستخدم معادلة التسارع الأساسية لتقدير القوة التي ضُربت بها كرة البيسبول خلال مباريات نهائيات كأس العالم. كان أمرًا ممتعًا، وكذلك كانت لِذَاتِ تلك المعادلة العديد من التطبيقات المهمة على المستوى الاجتماعي.

وما إن التحقتُ بالجامعة حتى وقعت في غرام قوانين الاحتمالات؛ وقد كان هذا أيضًا لأنها كانت تقدِّم لي فهمًا عميقًا لمواقف حياتية واقعية. ولقد أدركت متأخرًا أن الرياضيات لم تكن هي ما يصدني عن حسابات التفاضل والتكامل، وإنما كان هذا لأن أحدًا ما لم يكلف نفسه عناء أن يبين لي ما الغاية منها. فإذا لم تكن ممن يفتتنون بأناقة المعادلات في حد ذاتها – وأنا بكل تأكيد لست منهم – فلن تجد في هذه الحسابات إلا عملًا رتيبًا آليًّا، على الأقل بالطريقة التي دُرِّسَت لي بها.

ينقلنا هذا إلى الإحصاء (والذي يتضمن في هذا الكتاب قوانين الاحتمالات). إنني أحب الإحصاء لأنه قادر على تفسير كل شيء من اختبارات الحمض النووي إلى حماقة شراء تذاكر اليانصيب. والإحصاء يمكن أن يساعدنا في التعرف على العوامل المرتبطة بأمراض كالسرطان ومرض القلب، ويمكن أن يساعدنا على رصد الغش في الاختبارات الموحدة. بل إن الإحصاء يمكن أن يساعدك على الربح في برامج المسابقات التلفزيونية. وقت طفولتي كان هناك برنامج تلفزيوني شهير يسمى: لنعقد صفقة، يقدمه المذيع الشهير كذلك مونتي هول. في نهاية كل حلقة من حلقات البرنامج كان المتسابق الذي وصل للنهاية بنجاح يقف بجوار مونتي أمام ثلاثة أبواب مرقمة على نحو: الباب ١ والباب ٢ والباب ٣. ثم يوضح مونتي هول للمتسابق أن خلف واحد من هذه الأبواب الثلاثة تكمن جائزة قيِّمة – كسيارة جديدة أو ما نحوها – وخلف كلٍّ من البابين الآخرين تتوارى عنزة. تتلخص الفكرة مباشرة في اختيار المتسابق لأحد الأبواب ثم الظفر بما وراءه.

بينما كان كل لاعب يقف مواجهًا الأبواب مع مونتي هول يجد نفسه أمام فرصة من ثلاث فرص لربح الجائزة القيِّمة. لكن البرنامج كان يعتمد حيلة لم تفتأ تسعد علماء الإحصاء (وتربك غيرهم). فبعد أن يختار اللاعب أحد الأبواب يقوم مونتي هول بفتح واحد من البابين المتبقيين ودائمًا ما تظهر خلفه العنزة. لمزيد من تبيان هذا المثال لنفترض أن المتسابق اختار الباب رقم ١، حينها سيفتح مونتي الباب رقم ٣ فتظهر وراءه العنزة. هنا يظل البابان ١ و٢ مغلقَين. إذا كانت الجائزة خلف الباب ١ سوف يفوز المتسابق، أما إذا كانت خلف الباب ٢ فسيخسر. لكن في تلك اللحظة تتزايد الإثارة ويبلغ التشويق مبلغه؛ إذ يستدير مونتي للمتسابق ويسأله عما إذا كان يرغب في تغيير رأيه واختيار باب آخر (أن يتحول في هذه الحالة عن الباب ١ إلى الباب ٢). تَذكَّر أن كلا البابين مغلق وأن المعلومة الوحيدة الجديدة هنا هي أن عنزة من العنزتين ظهرت خلف واحد من البابين اللذين لم يخترهما.

هل يجدر به أن يبدل اختياره؟

الإجابة نعم. لماذا؟ سيأتي هذا في الفصل ½٥.

إن التناقض الذي يسم الإحصاء يتمثل في كونه مندمجًا في كل مناحي الحياة – من معدلات ضربات الكرة في مباريات الكريكيت إلى استطلاعات الرأي عن المرشحين الرئاسيين – إلا أن منهجه ذاته موصوم بسمعة سيئة لكونه صعبًا غير جاذب للاهتمام. فالعديد من الكتب المتخصصة في الإحصاء والفصول الدراسية التي تدرسه تتخمها حسابات رياضية ورطانة غير مفهومة. صدقني عندما أقول لك إن التفاصيل الفنية في هذا الأمر جوهرية (ومهمة)، لكنك إذا لم تكن من أصحاب التفكير المنطقي فستجدها مجرد كلمات يونانية. وقد لا تهتم أنت حتى بذلك التفكير المنطقي إذا لم تجد سببًا يقنعك بتعلمها. كل فصل من فصول هذا الكتاب يعدك بالإجابة على ذاك السؤال الأساسي الذي سألته (ولم أجد له إجابة وقتها) لمعلمتي في المدرسة الثانوية: ما الغاية من هذا؟

يتمحور هذا الكتاب حول التفكير المنطقي. وهو كتاب فقير في العمليات الحسابية والمعادلات الرياضية والرسوم البيانية، وأعدك أيها القارئ أنني إذا ما استخدمتها فسيكون هذا لخدمة أغراض واضحة منيرة. لكنك، في ذات الوقت، ستجده غنيًّا بأمثلة تهدف لإقناعك بوجود أسباب مهمة لتعلم هذه الأمور. إن الإحصاء قادر على أن يكون جاذبًا للاهتمام، وهو ليس صعبًا لهذه الدرجة في معظمه.

ولدت فكرة هذا الكتاب لديَّ بعد فترة قصيرة من واقعتي المؤسفة مع السيدة سميث في صف التفاضل والتكامل. كنت قد ارتدت كلية الدراسات العليا لدراسة الاقتصاد والسياسة العامة. وقبل حتى أن يبدأ البرنامج، تم توزيعي (وهو ليس بالأمر المستغرب) على «معسكر الرياضيات» مع بضعة من زملاء صفِّي بهدف تهيئتنا لما سنخضع له من قواعد كمية صارمة. وفي هذا المعسكر قضينا ثلاثة أسابيع ندرس الرياضيات يوميًّا في قاعة دراسة لا نوافذ لها تقع في قلب أحد الأقبية (صدقًا).

وفي واحد من تلك الأيام وجدتني أمام كشف منير يتعلق بمسيرتي المهنية. كان معلمنا يحاول أن يشرح لنا الظروف التي عند توافرها يتحول مجموع سلسلة غير محدودة إلى عدد محدد. فلتُعِرْني انتباهك هنا لدقيقة لأن المفهوم سيتضح هنا. (لعلك الآن تحس بما أحسسته عندما كنت في قاعة الدراسة العديمة النوافذ تلك). إن المتوالية اللانهائية هي عبارة عن تسلسل رقمي مستمر للأبد على غرار $1 + ½ + ¼ + ⅛$ ... وتعني تلك النقاط الثلاث أن هذا التسلسل يستمر إلى ما لا نهاية.

كان ذلك هو الجزء الذي صعب علينا استيعابه. كان معلمنا يحاول إقناعنا عن طريق دليل نسيته منذ وقت طويل بأن أرقامًا متسلسلة للأبد يمكن أن تتجمع (تقريبيًّا) فتصير رقمًا محددًا. لم يتمكن ويل وورشاور، زميلي في الصف، من فهم الأمر أبدًا رغم الدليل الرياضي المذهل. (وللأمانة أقول إنني نفسي كنت متشككًا في الأمر). فكيف لشيء لا محدود أن يتجمع ليصير شيئًا محدودًا؟

ثم حدث أن هبط عليَّ الإلهام، أو بالأحرى هبط عليَّ التفكير المنطقي، فأدركت ما يحاول المعلم أن يقول، فاستدرت إلى ويل وأخبرته بما اعتمل في ذهني لتوّه، فقلت له: تخيل أنك واقف على بُعد قدمين بالضبط من الجدار.

ثم تحركتَ نحوه قاطعًا نصف تلك المسافة (أي قدمًا واحدة بالضبط) فلا يصير يفصلك عن الجدار سوى قدم واحدة فقط.

ثم تتحرك مرة أخرى نحوه قاطعًا نصف المسافة (أي ست بوصات أو نصف قدم). ثم من على بُعد ست بوصات تكرر ما فعلت (أي تتحرك ثلاث بوصات أو ربع قدم). ثم تكررها (تتحرك لبوصة ونصف أو لثُمن قدم). وهلم جرًّا.

إنك تظل تقترب جدًّا من الجدار بشكل تدريجي. (فمثلًا عندما تكون على بُعد جزء واحد من ١٠٢٤ جزء من البوصة من الجدار سوف تتحرك بعدها لنصف المسافة أي جزء من ٢٠٤٨ من البوصة). لكنك لن تصطدم بالجدار أبدًا لأنك، كما تقتضي تلك العملية، لا تقطع سوى نصف المسافة الباقية التي تفصلك عن الجدار. أو بعبارة أخرى أقول إنك تظل تقترب من الجدار بشكل لا نهائي ولا تصطدم به أبدًا. وإذا ما قسنا حركاتك بالأقدام يمكن التعبير عن تلك المتوالية بالشكل التالي $1 + ½ + ¼ + ⅛$

هنا تكمن الفكرة: فرغم كونك تواصل التحرك للأبد – حيث تقطع في كل حركة نصف المسافة التي تفصلك عن الجدار – فإن مجمل المسافة التي ستمشيها لن يتجاوز القدمين بأية حال، وهي المسافة الأولية التي بدأتَ منها متجِهًا نحو الجدار. من الناحية الرياضية سيقدَّر مجمل المسافة التي ستقطعها بما يقارب القدمين، وهو الأمر الذي ستستفيد منه الناحية الحسابية كذلك. فالرياضيون سيقولون إن مجموع تلك المتوالية اللامتناهية قدم واحدة + ½ قدم + ¼ قدم + ⅛ قدم ... سيتحول إلى قدمين، وهي الفكرة التي ما فتئ معلمنا يحاول إفهامنا إياها طوال ذلك اليوم.

الفكرة هنا هو أنني أقنعت ويل وأقنعت نفسي بهذا الأمر. إنني لا أذكر الحسابات الرياضية التي تبرهن على أن مجموع متوالية لا نهائية يمكن أن يحوَّل إلى عدد محدود، لكنني أستطيع أن أحصل على تلك الحسابات بكل سهولة من شبكة الإنترنت. وعندما أفعل ذلك سوف أجدها منطقية على الأرجح. لقد عرفت من خبرتي أن التفكير المنطقي يجعل الحسابات الرياضية وغيرها من التفاصيل الفنية أكثر قابلية للفهم – لكن العكس ليس صحيحًا بالضرورة.

إن الغرض من هذا الكتاب هو جعل أكثرِ المفاهيم الإحصائية أهمية أكثرَ منطقية وسهولة، لا للمضطرين لدراستها في قاعات دراسة عديمة النوافذ فقط، وإنما كذلك لكلِّ مَن هم مهتمون بما للأرقام والبيانات من سطوة طاغية.

الآن بعد أن أثبتنا أن أدوات الإحصاء الأساسية أقل منطقية وإتاحة مما يجب أن تكون عليه، سوف أعرض وجهة نظر مناقضة، ألا وهي أن الإحصاء أكثر إتاحة لدرجة أن أي شخص يملك بعض البيانات وجهاز كمبيوتر قادر على إنجاز إجراءات إحصائية معقدة بكبسات أزرار قليلة. لكن المشكلة هنا أنه إذا كانت البيانات شحيحة أو إذا ما استخدمت الأساليب الإحصائية بشكل خطأ، قد تأتي النتائج مضللة، بل إنها قد تكون خطيرة. فلنتأمل هذا الخبر الافتراضي الإلكتروني التالي: إن مَن يحصلون على فترات استراحة قصيرة خلال عملهم هم أكثر عرضة للوفاة جراء مرض السرطان. تخيل هذا العنوان يقفز أمام ناظريك بينما تتصفح الشبكة العنكبوتية. طبقًا لدراسة تبدو مهمة أجريت على ٣٦٠٠٠ من موظفي المكاتب (وهو ما يعني قدرًا هائلًا من البيانات) هؤلاء الموظفون الذين أقروا بأنهم يتركون مكاتبهم في كل يوم عمل لفترة استراحة تقدَّر بعشر دقائق هم أكثر عرضة للإصابة بمرض السرطان خلال الخمس سنوات التالية بنسبة أعلى من نظرائهم الذين لا يتركون مكاتبهم بنسبة ٤١ بالمائة. من هذا يتضح أن علينا أن نتخذ إجراءً وفق هذه النتيجة، قد يتمثل هذا الإجراء في إطلاق حملة توعوية قومية تهدف إلى منع فترات الاستراحة القصيرة أثناء أوقات العمل.

أو قد نجد أنفسنا بحاجة إلى أن نفكر فيما يفعل أولئك الموظفون خلال الدقائق العشر تلك. من خبرتي العملية أظن أن العديد من أولئك الموظفين الذين ثبت أنهم يغادرون مكاتبهم يتجمعون حول مدخل البناية يدخنون السجائر (مسببين سحابة من الدخان يتعين علينا أن نخترقها عندما ندخل البناية أو نخرج منها). إنني سوف أتجرأ وأقول إن السجائر، لا فترات الاستراحة، هي التي تسبب السرطان للموظفين. إنني قد اختلقت هذا المثال قاصدًا أن يبدو بهذه السخافة، لكنني أؤكد لك أن الكثير من الأمثلة الإحصائية الواقعية في الحياة تتسم بسخف فاحش يضارع مثالي المختلق السابق.

إن الإحصاء أشبه بسلاح متفوق جدًّا ذي منفعة كبيرة إذا ما استُخدم بالشكل الصحيح، وكارثي إذا وقع في الأيدي الخطأ. لن يجعلك هذا الكتاب خبيرًا في الإحصاء، وإنما سيغرس فيك اهتمامًا واحترامًا لذلك المجال كي لا تفعل به ما يفعله من يسيء استخدام السلاح فيردي أحدهم برصاصة في رأسه.

ليس هذا بكتاب أكاديمي، وهو الأمر الذي جعلني أكثر حرية من جانب المواضيع التي غطَّيتُها والطرائق التي انتهجتُها في الشرح. ولقد صُمم هذا الكتاب للتعريف بالمفاهيم الإحصائية بشكل أكثر ارتباطًا بالحياة اليومية. كيف يتسنى للعلماء أن يستنتجوا أن شيئًا ما قد يسبب السرطان؟ ما آلية عمل استطلاعات الرأي؟ (وكيف تنحو منحًا خطأً؟) من «يكذب بالإحصائيات»؟ وكيف يفعلها؟ كيف تستخدم شركة بطاقة الائتمان التي تملكها البيانات المتعلقة بمشترياتك في التنبؤ بأنك قد تتخلَّف عن أحد أقساطك؟ (نعم هم قادرون على فعل هذا).

إذا ما أردت أن تفهم الأرقام التي تكمن خلف الأخبار وأن تدرك تلك القوة الخارقة (والمتنامية) التي تتمتع بها البيانات، فتلك هي الأمور التي تحتاج لأن تعرفها. وأخيرًا إنني آمل أن أقنعك بتلك الملاحظة التي كان الرياضي والكاتب السويدي آندريس دانكلز أول من أشار إليها، والتي مفادها أنه من السهل أن تكذب بالإحصاء، لكن من الصعب أن تقول الحقيقة من دونه.

لكنني آمل فيما هو أكبر من ذلك، فإنني أظنك قادرًا على أن تستمتع بالإحصاء. فالأفكار الكامنة وراءه شائقة وسهلة إلى حد مذهل، والسبيل إلى ذلك هو فصل الأفكار المهمة عن التفاصيل التقنية الغامضة التي قد تحول بينك وبينها. ذاك هو الإحصاء عاريًا.

شكر وتقدير

إن هذا الكتاب هو استلهام لروح كتاب آخر يسبقه من كلاسيكيات دار دبليو. دبليو. نورتون للنشر بعنوان How to Lie with Statistics لمؤلفه داريل هاف، وهو الكتاب الذي صدر في خمسينيات القرن الماضي، وبِيعَ منه ما يربو عن المليون نسخة. وقد كُتب ذاك الكتاب، مثل كتابنا هذا، بهدف إزالة ما يحيط الإحصاء من غموض، وكذلك لإقناع القارئ العادي أن ما لا يدركه عن تلك الأرقام التي تتوارى خلف عناوين الأخبار هو أنها قادرة على إيذائه. إنني آمل أن أكون قد وفيت كتاب السيد هاف الكلاسيكي حقه، فإنني لو عاصرت زمانه قبل خمسين عامًا من الآن لسَرَّني أن أبيع أكثر من مليون نسخة من كتابي.

إن امتناني تجاه دار دبليو. دبليو. نورتون للنشر وتجاه دريك ماكفيلي دائم لا ينقطع؛ لأنهما مكَّناني من أن أؤلف كتبًا تناقش مواضيع مهمة بطريقة يفهمها القارئ المعاصر. ولقد ظل دريك صديقًا عزيزًا داعمًا لي لأكثر من عقد من الزمان.

وجيف شريف أحد رجال دبليو. دبليو. نورتون الذي أخذ يدفع عملية إنجاز الكتاب دفعًا حتى تحققت في النهاية. قد ينخدع من يقابل جيف لأول مرة، فيظنه أكثر لطافة وطيبة من أن يقدر على فرض الالتزام بالمواعيد النهائية لمراحل إنجاز الكتاب، لكن هذا الانطباع الأوَّلي غير صحيح على الإطلاق. نعم هو لطيف جدًّا، لكن طريقته الرقيقة في الحث على الإنجاز تؤتي ثمارها (فمثلًا كلمات الشكر والتقدير هذه لا بد أن تُسلَّم إليه صباح الغد). وإنني أقدِّر جيدًا قيمة وجود مدير المهام العطوف ذلك ليُسيِّر لي أموري.

القدر الأعظم من امتناني أدين به لأولئك الرجال والنساء الكثيرين الذين قاموا بتلك الأبحاث والتحليلات المهمة التي عرضتُها في هذا الكتاب. فإنني لست بإحصائي ولا باحث. ما أنا إلا مترجم لتلك الأعمال المهمة والمشوقة التي أنجزها أولئك الأشخاص. آمل أن أكون قد أوصلت من خلال هذا الكتاب فكرةً مُفادُها أن العمل البحثي الجيد والتحليل المتماسك من شأنهما أن يجعلانا أكثر صحة وثراءً وأمنًا ومعرفة.

وأخص بالشكر أستاذ الاقتصاد بجامعة برنستون آلان كروجر لمساهماته الجليلة التي جاءت في صورة أبحاث متسعة النطاق، تدرجت مواضيعها ما بين البحث في جذور الإرهاب إلى عوائد التعليم العالي. (لقد كانت النتائج التي خلص إليها في هذين الموضوعين غير متوقعة بشكل يبعث على السرور). لكن أهم من ذلك (بالنسبة إليَّ) أن آلان كان واحدًا من أساتذة الإحصاء الذين دَرَستُ عليهم في مرحلة الدراسات العليا، ولطالما بهرتني قدرته على الموازنة بين البحث العلمي والتدريس والخدمة العامة.

وكذلك جيم سالي وجيف جروجر وباتي آندرسون وآرثر مينتز، الذين قرءوا مسودات الكتاب الأولية وزودوني بالكثير من الاقتراحات المفيدة. أشكركم لأنكم حميتموني من نفسي. وكان فرانك نيوبورت من مؤسسة جالوب ومايك كاجاي الصحفي بجريدة النيويورك تايمز، من الطيبة أن أنفقا بعض الوقت في إفهامي دقائقَ منهجيةِ عمليةِ استطلاعِ الرأيِّ. ورغم كل جهودهم جاءت تلك الأخطاء الباقية مني.

لقد كانت كاتي ويد مساعدة بحثية «لا تكل ولا تمل» (لطالما وددت أن أستخدم ذلك التعبير، وأخيرًا وجدت السياق المثالي له). ولقد جاءت معظم الحكايات الطريفة والأمثلة التي أضاءت الكثير من المفاهيم الواردة في الكتاب من كاتي. فمن دون كاتي لا أمثلة فكاهية.

ولقد قام الكثير من القراء الذين قرءوا النسخة ذات الغلاف الصلب من الكتاب، بإرسال العديد من التوصيات والملاحظات إليَّ. شكر خاص لجاي ديفور وإدوارد لنشتز اللذين قدَّما إليَّ بعض الحيل المفيدة وأشارا إلى بعض النقاط الفنية المهمة.

لطالما رغبت في أن أؤلف الكتب مذ كنت في المدرسة الابتدائية. تينا بينيت، وكيلَتي، هي الشخص الذي مكَّنني من أن أفعل ذلك، بل إنها مكَّنتني كذلك من أن أتخذها مهنة. إن تينا هي تجسيد لأفضل ما في صناعة النشر، فهي تجد سعادتها في إنجاز العمل، بينما تقوم في ذات الوقت بالاعتناء بمصالح عملائها.

وأخيرًا تستحق أسرتي الثناء لتحمُّلها إياي طوال مرحلة إنتاجي لهذا الكتاب. (لقد كنت أعلق المواعيد النهائية لتسليم كل فصل من فصول الكتاب على باب الثلاجة). وهناك دلائل على أنني كنت أصير أكثر عصبية بنسبة ٣١ بالمائة وأكثر إرهاقًا بنسبة ٢٣ بالمائة عندما كنت أدنو من أحد تلك المواعيد النهائية الأساسية (أو لا أفي به). وزوجتي، ليا، هي أول وأفضل وأهم محرر لكل ما أكتب. أشكركِ على ذلك وأشكركِ لأنكِ ظللتِ شريكًا ذكيًّا داعمًا مرحًا في كل مساعيَّ الأخرى.

أُهدي هذا الكتاب إلى ابنتي الكبرى، كاترينا. فمن الصعب تصديق أن تلك الطفلة التي كانت لا تزال في مهدها وقت أَلَّفتُ كتاب «الاقتصاد عاريًا»، صارت الآن تقرأ فصولًا وتسجل ملاحظات مهمة عليها. إنكِ يا كاترينا حلمُ والدَيك، ومعك صوفي وسي جيه اللذان هما أيضًا سيقومان قريبًا بقراءة فصول ومخطوطات كذلك.

الفصل الأول

ما المقصد؟

إنني قد لاحظت ظاهرة محيرة؛ وهي أن التلاميذ دائمًا ما يشتكون من كون علم الإحصاء علمًا مُلْغِزًا وغير ذي أهمية، في حين أنهم ذاتهم بعد أن يغادروا صفوف الدراسة ويجتمعوا في حصص الغداء ينخرطون في نقاشات ماتعة حول متوسطات مرات ضرب الكرة في مباريات البيسبول (خلال فصل الصيف)، أو معامل تبريد الرياح (خلال فصل الشتاء)، أو متوسطات نقاط الدرجات الدراسية (في كل الفصول). إنهم يدركون أن «معدلات تمرير الكرة» في البطولة الوطنية لكرة القدم – التي هي عبارة عن إحصائية تختزل أداء أحد اللاعبين في مركز خط الوسط لرقم مفرد – هي معيار معيب عشوائي لأداء واحد من لاعبي هذا المركز. وهذه البيانات ذاتها (التي تتمثل في كلٍّ من معدلات اكتمال الهجمات، ومتوسط الياردات التي تُقطع في كل محاولة هجمة، ومتوسط التمريرات التي تتم في منطقة إحراز الأهداف في كل محاولة هجمة، ومعدل التدخلات الاعتراضية) يمكن أن تُجمع بوسائل أخرى، كتركيز الاهتمام أو تقليله عن واحد من تلك المدخلات بهدف توليد معيار مختلف للأداء، لكنه في ذات الوقت لا يقل عن سابقه مصداقية. لكنَّ أيَّ مُشاهد لمباريات كرة القدم، رغم كل ما سبق، يدرك كم هو من المفيد اختزال أداء لاعب خط الوسط في رقم واحد مفرد.

هل معدلات أداء لاعب خط الوسط مثالية؟ كلا، فهذه الإحصائيات لا تكاد تقدِّم طريقة واحدة «جديدة» لعمل أي شيء. هل تقدِّم معلومات مهمة بطريقة سهلة ميسرة؟ بالقطع، فهي أداة جيدة لعقد مقارنات سريعة بين أداءين لاثنين من لاعبي خط الوسط في يوم واحد. إنني من مشجعي فريق شيكاجو بيرز. وخلال منافسات المربع الذهبي لعام ٢٠١١ لاعَب البيرز فريق الباكرز ففاز الباكرز. إنني قادر على تحليل تلك المباراة بطرائق متعددة، من بينها كتابة صفحات وصفحات من تحليلات البيانات الخام. لكنْ إليك هذا التحليل الموجز. بلغ معدل تمريرات لاعب خط الوسط في فريق البيرز، جاي كاتلر ٣١٫٨. بينما بلغ معدل تمريرات آرون رودجرز لاعب خط الوسط لدى فريق جرين باي ٥٥٫٤. بالمثل يمكننا أن نقارن أداء جاي كاتلر في مباراة أخرى ضد فريق الجرين

باي في جولة مبكرة من الدوري حين بلغ معدل تمريراته ٨٥,٦. هذه المقارنات تُنبئك بالكثير مما تريد معرفته كي تدرك لماذا هَزم البيرز الباكرز في أول الموسم ثم انهزم أمامه في منافسات المربع الذهبي.

ذاك موجز مفيد جدًّا يبين ما حدث في الملعب. هل بسَّط الأمرَ؟ نعم قد فعل، وهو الأمر الذي يمثل ما تحوزه أي إحصائية وصفية من قوة وما يعتريها من ضعف في ذات الوقت. فواحد من تلك الأرقام يُنبئنا بأن آرون رودجرز قد تفوَّق على جاي كاتلر في مباراة المربع الذهبي التي خسرها البيرز. لكن في ذات الوقت لن يخبرك هذا الرقم إذا ما كان لاعب خط الوسط ذلك قد صادف حظًّا عاثرًا عندما مرر الكرة تمريرة مثالية لكن زميله المتلقي أفلتها، فاعترضها لاعب الفريق الآخر، أو إذا ما كان قد «انبرى مبادرًا» بهجمات خطيرة (بما أن كل الهجمات تتساوى في الإحصاء، سواء كانت هجمة خطيرة من الرمية الثالثة، أو لعبة لا جدوى منها في نهاية المباراة)، أو إذا ما كان الدفاع سيئًا جدًّا، أو أي شيء آخر.

لكن المثير للفضول هنا هو أن أولئك الأشخاص ذاتهم الذين ينخرطون باستمتاعٍ في مناقشة الإحصاء في سياق الرياضة أو حالة الطقس أو الدرجات الدراسية، يجزعون ويقلقون عندما يَشْرع أحد الباحثين في شرح مواضيع كمعامل جيني، الذي هو من أدوات الاقتصاد المعيارية لقياس التفاوتات في الدخل. سوف أشرح فيما يلي ما هو معامل جيني، لكن أهم ما علينا أن ندركه الآن هو أن معامل جيني لا يختلف عن معدلات تمرير الكرة. إنه أداة تساعد كثيرًا في اختزال المعلومات المعقدة إلى رقم مفرد. وهو بهذا يملك مَواطن القوى التي تتمتع بها معظم الإحصائيات الوصفية، فهو قادر على تقديم وسيلة سهلة للمقارنة بين توزيع الدخل في دولتين مختلفتين أو في دولة واحدة في فترتين زمنيتين متفرقتين.

يقيس معامل جيني كيف يتم تقاسم الثروة (أو الدخل) في داخل الدولة على مقياس يتدرج من صفر إلى واحد. يمكن حساب تلك الإحصائية بالنسبة إلى الثروة أو إلى الدخل السنوي، ويمكن أن تُحسب على مستوًى فردي أو عائلي. (يمكن أن تتسق تلك الإحصائيات بعضها مع بعض لكنها لن تتطابق أبدًا). ومعامل جيني مثل معدلات تمرير الكرة لا يحمل معنًى داخليًّا ينبع من ذاته، وإنما هو مجرد أداة للمقارنة. وبالنسبة إلى دولة تتمتع كل عائلة منها بذات القدر من الثروة، يُقدَّر معامل جيني لديها بصفر. وفي المقابل يُقدَّر معامل جيني لدى الدولة التي استولت فيها عائلةٌ واحدة على كل الثروة بواحد. وكما قد استنتجت على الأرجح، كلما اقتربت دولةٌ ما من قيمة واحد بمعامل جيني، كانت الثروة فيها موزعة بغير تَساوٍ. يُقدَّر معامل جيني في الولايات المتحدة بـ ٠,٤٥ طبقًا لوكالة المخابرات المركزية الأمريكية (والتي هي بالمناسبة جامع بيانات ممتاز)[١]، ماذا يعني هذا إذن؟

ما إن يوضع هذا الرقم في سياقه حتى يخرج علينا بالكثير من المعلومات، فمثلًا يبلغ معامل جيني لدى السويد ٠,٢٣، بينما يبلغ في كندا ٠,٣٢، ويبلغ في الصين ٠,٤٢،

وفي البرازيل ٠,٥٤، وفي جنوب أفريقيا *٠,٦٥. وعندما نتأمل تلك الأرقام نبدأ في استشعار موقع الولايات المتحدة من بقية البلدان فيما يتعلق بتفاوتات الدخل. كما أننا نستطيع كذلك أن نقارن بين الوضع في فترات زمنية مختلفة. لقد قُدِّر معامل جيني في الولايات المتحدة عام ١٩٩٧ بـ ٠,٤١، ثم نما إلى ٠,٤٥ خلال العقد الذي تلا ذلك العام. (صدرت أحدث بيانات وكالة المخابرات المركزية الأمريكية عام ٢٠٠٧). وهذا يُنبئنا بصورة موضوعية أنه مع ازدياد ثراء الولايات المتحدة خلال تلك الفترة الزمنية، تزايَد التفاوت في توزُّع تلك الثروة. وكذلك يمكننا المقارنة بين التغيرات التي طرأت على معامل جيني في البلدان المختلفة في ذات الفترة الزمنية. فالتفاوت في الدخل لم يتغير في كندا خلال تلك الفترة الزمنية. أما السويد التي شهدت نموًّا اقتصاديًّا كبيرًا خلال العقدين الماضيين فقد هبط فيها معامل جيني من ٠,٢٥ عام ١٩٩٢، إلى ٠,٢٣ عام ٢٠٠٥؛ وهو ما يعني أنه مع ازدياد ثراء السويد تَوزُّع الدخل فيها بشكل أكثر تساويًا.

لكن هل يعتبر معامل جيني المعيار المثالي لقياس التفاوتات؟ بالقطع لا، تمامًا كما أن معدل تمرير الكرة ليس بالمقياس المثالي لأداء لاعب خط الوسط في كرة القدم الأمريكية. لكنه بكل تأكيد يمنحنا معلومات قيِّمة عن ظواهر اجتماعية مهمة بشكل تقليدي.

إننا الآن قد تقدمنا ببطء نحو الإجابة عن ذلك السؤال الذي عَنْوَنَّا به هذا الفصل، ألا وهو: ما المقصد؟ المقصد هنا هو أن الإحصاء يساعدنا على تحليل البيانات، والتي هي مجرد إعادة تسمية منمَّقة للمعلومات. أحيانًا ما تكون البيانات أمرًا تافهًا وسط أمور أكثر عظمة، كما هو الحال في الإحصائيات الرياضية. وأحيانًا نجدها تقدم لنا مناظير بصيرة لطبيعة الوجود الإنساني، كما هو الحال مع معامل جيني.

لكن كما في أي دعاية تلفزيونية جيدة، لا بد أن نقول هنا إن هذا ليس كل شيء. قال هال فاريان، كبير الاقتصاديين بموقع جوجل، لصحيفة النيويورك تايمز إن وظيفة المختص بالإحصاء ستصير «أكثر الوظائف جاذبية» خلال العقد القادم[٢]. سأكون أول من يقر بأن الاقتصاديين أحيانًا ما يتمتعون بخاصية «الجاذبية» تلك، لكن مع هذا دعنا نتأمل الأسئلة المتباينة التالية:

كيف يمكننا أن نضبط المدارس التي تغش في اختباراتها الموحدة؟

كيف لمنصة نيتفليكس للبث الرقمي أن تعرف نوعية الأفلام التي تحبها؟

كيف لنا أن نحدد المواد أو السلوكيات التي تسبب السرطان، مع كوننا لا نستطيع إجراء تجارب مسببة للسرطان على البشر؟

هل تفيد الصلاة لأجل مرضى الجراحات في علاجهم؟

---

* أحيانًا ما يتم ضرب رقم معامل جيني في ١٠٠ ليصير رقمًا صحيحًا، وفي تلك الحالة يبلغ معامل جيني للولايات المتحدة ٤٥.

هل هناك بالفعل فائدة اقتصادية من وراء الحصول على درجة علمية من إحدى الجامعات العريقة؟

ما الذي يسبب ارتفاع نسبة الإصابة بداء التوحد؟

يساعد الإحصاء في الإجابة على تلك التساؤلات (أو سيساعد في هذا عما قريب كما نأمل).

إن العالم لا يفتأ يُنتج المزيد والمزيد من البيانات ويزداد في هذا سرعة. لكن وكما تقول صحيفة النيويورك تايمز: «ما البيانات إلا المادة الخام للمعرفة»[٣]*. إن الإحصاء هو أقوى ما لدينا من أدوات لاستخدام المعلومات لأجل غاية ذات معنًى، سواء كانت تلك الغاية هي رصد لاعبي البيسبول الذين لم يَلقَوا حقهم من الشهرة، أو حتى مجازاة المعلمين بما يستحقون من رواتب. فيما يلي جولة سريعة نرى من خلالها كيف يضفي الإحصاء معنًى على البيانات الخام.

الوصف والمقارنة

إن نتيجة مباراة من مباريات البولينج هي إحصائية وصفية، وكذلك هو معدل ضرب الكرة في البيسبول. ومعظم مشجعي الرياضات الأمريكيين الذين تتعدى أعمارهم الخمسة أعوام، هم ملمون بالفعل بمجال الإحصاء الوصفي. فنحن نستخدم الأرقام في الرياضة وفي أي مجال آخر لتلخيص المعلومات. كم كان ميكي مانتل لاعب البيسبول لاعبًا جيدًا؟ لقد ضرب ٢٩٨ ضربة خلال مسيرته الرياضية. تلك المقولة بالنسبة إلى مشجع لرياضة البيسبول هي عبارة ذات معنًى، هي عبارة فذة إذا ما تدبَّرتَها؛ إذ تلخص مسيرة مهنية استمرت لثمانية عشر موسمًا[٤]. (لكنني أظنه أمرًا يبعث على بعض الكآبة أن تختصر الحياة المهنية لشخصٍ ما في رقم واحد). وبالطبع أدرك مشجعو البيسبول أن الإحصائيات الوصفية بخلاف معدلات ضرب الكرة من الأفضل أن تُستخدم في تلخيص قيمة اللاعب في الملعب.

إننا نقيِّم أداء التلاميذ الأكاديمي في المدرسة الثانوية وفي المرحلة الجامعية من خلال متوسط نقاط التقييم GPA. الحرف أ يعبر عن درجة من درجات التقييم، وعادة ما يعبر هذا الحرف عن أربع نقاط، بينما يعبر الحرف ب عن ثلاثة، ويعبر الحرف ج عن نقطتين،

* دائمًا كانت كلمة «data» الإنجليزية تعتبر صيغة جمع وتعني «البيانات». (كما في عبارة «البيانات مشجعة للغاية») ومفردها كلمة «datum»، والتي تشير إلى وحدة مفردة من وحدات البيانات كاستجابة شخص مفرد لسؤال واحد في استطلاع رأي. واستخدام كلمة «data» كاسْم في صيغة الجمع هو طريقة سريعة تشير بها لأي ممن يُجرون أبحاثًا جادة إلى أنك مُلمٌّ بالإحصاء. لكن مع هذا فإن الكثير من العلماء الثقات في مجال النحو، بالإضافة إلى الكثير من المطبوعات الصحفية على غرار النيويورك تايمز، يَقبلون استخدام تلك الكلمة بقصد الإفراد أو الجمع على السواء، كما اتضح في تلك الفقرة التي اقتبستها من صحيفة التايمز.

وهلم جرًّا. ومع بلوغ مرحلة التخرج عندما يتقدم تلاميذ المدارس الثانوية للالتحاق بالجامعة ويتقدم طلبة الجامعات للالتحاق بوظائف، يصير متوسط نقاط التقييم أداة مفيدة في تقييم إمكانياتهم الأكاديمية المنتظرة. فالطالب الذي أحرز ٣٫٧ نقاط هو – كما هو واضح – أقوى إمكانياتٍ من زميله ذي الـ ٢٫٥ نقطة. ذاك يجعل منها إحصائية وصفية رائعة. فهي سهلة الحساب سهلة الفهم، ومن السهل أن تكون مجالًا للمقارنة بين الطلبة.

لكنها ليست خالية من العيوب. فمتوسط نقاط التقييم لا يعكس بصدق صعوبة المقررات الدراسية التي درسها مختلف التلاميذ. فكيف لنا أن نقارن تلميذًا أحرز ٣٫٤ نقاط من دراسته لمقررات يسيرة، بتلميذ آخر أحرز ٢٫٩ نقطة من دراسة حساب التفاضل والتكامل والفيزياء وغيرها من المقررات الصعبة؟ حاولت المدرسة الثانوية التي درست فيها أن تحل تلك المشكلة عن طريق منح درجات إضافية للمقررات الصعبة، فقد كان حرف أ في صف «المميزين» يساوي خمس نقاط، لا أربعة كما في غيره من الصفوف. لكن هذا أدى إلى مشكلات من نوعية أخرى. سرعان ما أدركت أمي هذا الانحراف الذي سببه تعديل نقاط التقييم ذاك. فلِكَوني تلميذًا درست الكثير من مقررات المميزين، كان أيُّ تقدير أ أناله في مقرر ليس من ضمن مقررات المميزين – كمقرر الرياضة البدنية أو مقرر الصحة العامة – سوف يهبط بمتوسط تقييمي رغم أنه من المستحيل أن أنال في تلك المواد ما يزيد عن التقدير أ. ونتيجة لهذا مَنَعني أبواي من أن ألتحق بصف تعليم القيادة في المدرسة الثانوية، رغم أنني كنت سأُبلي فيه بلاءً حسنًا، خشية أن يقلل هذا من فرصي في الالتحاق بجامعة عريقة وكتابة تلك الكتب ذائعة الصيت. وبدلًا من هذا أنفقا المال ليُلحقاني بمدرسة ليلية لتعليم القيادة خلال الفترة الصيفية.

هل كان ذاك ضرب من الجنون؟ نعم كان كذلك. لكن إحدى الفِكَر الأساسية في هذا الكتاب تقول بأن المبالغة في الاعتماد على أي إحصائية وصفية يمكن أن تؤدي إلى نتائج مضللة أو تتسبب في سلوكيات غير محمودة. كانت الجملة التي كتبتها أولًا للتعبير عن هذا المعنى «إحصائية وصفية مبالغ في تبسيطها»، لكنني حذفت عبارة «مبالغ في تبسيطها» لأنها بدت من قبيل الإسهاب غير الضروري. فما وُجدت الإحصائيات الوصفية إلا بغرض التبسيط، وهو الأمر الذي ينطوي على التخلص من الدقائق والتفاصيل. وكلُّ مَن يعمل بالأرقام يحتاج لأن يدرك ذلك المعنى.

الاستدلال

كم من المشردين يعيشون في شوارع شيكاجو؟ ما معدل ممارسة المتزوجين للجنس؟ قد يبدو لك أن هذين سؤالان مختلفان نوعيًّا للغاية، لكن الواقع أن كليهما يمكن

الإجابة عنه (بطريقة ليست مثالية) من خلال استخدام أدوات إحصائية أولية. فواحدة من وظائف الإحصاء الرئيسية هي استخدام البيانات المتاحة للخروج بتنبؤ قائم على معلومات يجيب عن الأسئلة ذات النطاق الأكثر اتساعًا، والتي لا نملك عنها المعلومات الكاملة. للاختصار نقول إننا نستطيع استخدام البيانات التي نحصل عليها من «العالم المعروف»، كي نقوم باستدلالات قائمة على معلومات عن «العالم المجهول».

لنبدأ بالسؤال الذي يتناول المشردين. إنه من المكلف والعسير من الناحية اللوجيستية أن نقوم بِعدٍّ كل المشردين الموجودين في منطقة حضرية. لكنه من المهم أن يكون لدينا تقديرًا عدديًّا لعددهم بغرض توفير الخدمات الاجتماعية لهم، وتحصيل العوائد المستحقة للولاية وللحكومة الفدرالية منهم، وكذلك كي يحظوا بتمثيل برلماني. واحد من الإجراءات الإحصائية المهمة يتمثل في أخذ العينات، والذي هو عملية جمع البيانات في نطاق مساحة صغيرة – كبضعٍ من المناطق المحدودة النطاق – ومن ثم استخدام تلك البيانات لإطلاق أحكام مبنية على معلومات، أو إجراء استدلال عن كامل مجتمع المشردين الذين يقطنون المدينة. إنَّ أخذ العينات يتطلب موارد أقل بكثير من تلك التي يتطلبها عد جميع المشردين في المدينة، وإذا ما أُجري بالصورة الصحيحة لن يقل عنه دقة.

تعتبر استطلاعات الرأي السياسية واحدة من صور أخذ العينات. فخلالها تحاول إحدى المؤسسات البحثية جمع العينات من أُسر تكون بشكلٍ ما الأكثر تمثيلًا لديموغرافية أكبر، لتسألها عن وجهة نظرها حول قضية معينة أو مرشح سياسي معين. وكما هو واضح، فإن هذه وسيلة أسرع وأقل تكلفة من محاولة التواصل مع كل أسرة في الولاية أو الدولة. ترى مؤسسة جالوب للأبحاث واستطلاعات الرأي أن استطلاعَ رأيٍ سليمًا من الناحية المنهجية يستقصي رأي ألف أسرة، سوف يخرج بذات النتائج تقريبًا التي ستخرج بها محاولة التواصل مع كل أسرة في أمريكا.

وبذات الطريقة عرفنا الوتيرة التي يمارس بها الأمريكيون الجنس، ومع من يمارسونه، وبأي نوع يمارسونه. ففي منتصف التسعينيات من القرن الماضي، قام مركز أبحاث الرأي العام بجامعة شيكاجو بدراسة طموحة لأقصى حد تبحث في السلوك الجنسي للأمريكيين. جاءت نتائج الدراسة قائمة على استقصاءات مفصلة أُجريت بشكل شخصي على عينة ممثلة للبالغين الأمريكيين. إذا واصلتَ القراءة فسوف تعرف من الفصل العاشر ما خرجَت به الدراسة. كم من كتب الإحصاء تَعِدك بشيء كهذا؟

تقييم المخاطر وغيرها من الأحداث المرتبطة بالاحتمالات

تربح صالاتُ لعب القمار المالَ على المدى الطويل. لكن هذا لا يعني أنَّ رِبحها محدد الموعد. فمع انطلاق الصفير وقرع الأجراس يربح بعض المقامرين البارعين آلاف الدولارات.

إن صناعة القمار بأسرها تقوم على ألعاب الحظ؛ بمعنى أن ناتج أي رمية نرد أو قلب بطاقة لعب هو أمر غير مؤكد. لكن في الوقت ذاته تكون احتمالات وقوع تلك الأحداث المتعلقة بألعاب الحظ تلك – كسحب ورقتين رابحتين مجموعهما ٢١ في لعبة البلاك جاك، أو استقرار عجلة الحظ الدوارة على اللون الأحمر – معروفة جدًّا. عندما تخدم الاحتمالات مصلحة صالات لعب القمار (كما تفعل دائمًا)، يتنامى لدينا يقين بأن مكاسب الصالة «أو المنزل كما هو اسمها أثناء اللعب» سوف تستمر، بينما يتزايد عدد المقامرين رغم انطلاق الصفير وقرع الأجراس.

اتضح أن تلك ظاهرة قوية تسود حتى خارج صالات لعب القمار. فالعديد من الشركات يتعين عليها أن تقيِّم المخاطر المتصلة بنتائج أعمالهم المتنوعة المتقلبة. فتلك الشركات لن تستطيع إبعاد الخطر كلية، تمامًا كما لا تستطيع صالات القمار أن تضمن الخسارة الدائمة لكل لعبة بلاك جاك تلعبها أنت. لكن أي شركة تواجه عدمَ يقينٍ من نتائجها يمكن أن تقلل تلك المخاطر المتنوعة من خلال هندسة عمليات معينة تؤدي إلى تقليل حدوث تقلبات في الاحتمالات، سواء كان هذا في صورة كوارث بيئية أو منتجات تجارية معطوبة. فمؤسسات وول ستريت تقوم بتقييم المخاطر المحدقة بمحافظها المالية من خلال سيناريوهات عديدة، كلٌّ منها يقيم من خلال احتمالاته. لقد اندلعت الأزمة الاقتصادية التي حدثت عام ٢٠٠٨، وتسارعت وتيرتها من جراء سلسلة من الأحداث السوقية التي ظُنت في البداية بعيدة الاحتمالية، بالضبط كما لو أن لاعب قمار ظل طوال الليلة يسحب الورقة الرابحة في لعبة البلاك جاك. لاحقًا سأشرح كيف أنني أجد نماذج وول ستريت تلك معيبة، وأن البيانات التي تعتمد عليها في تقييم المخاطر المتوقعة محدودةٌ أكثر من اللازم، لكن المقصد هنا هو أن أي نموذج يتعامل مع تلك المخاطر لا بد أن يرتكز على الاحتمالات.

وعندما يتعذر على الأفراد والمؤسسات أن يتجنبوا تلك المخاطر غير المقبولة، يتعين عليهم أن يطلبوا الحماية بسبل أخرى. فصناعة التأمين بأسرها تقوم على تقاضي الأموال من العملاء في مقابل حمايتهم من تقلبات الأحداث، على غرار حوادث السيارات أو احتراق المنازل. وصناعة التأمين تلك لا تجني مكاسبها من منع تلك الأحداث؛ فما زالت السيارات تتصادم والمنازل تحترق حتى يومنا هذا. بل وأحيانًا تصطدم السيارات بمنازل فتجعلها تحترق. وإنما تجني شركات التأمين مكاسبها بأن تتقاضى من العملاء أقساط تأمين هي أكثر مما يكفي لتغطية تكاليف حوادث السيارات وحرائق المنازل. (تقوم شركات التأمين أيضًا بمحاولة تقليل التكاليف المتوقعة من خلال تشجيع القيادة الآمنة، ووضع الأسْيِجة حول بِرَك السباحة، وتركيب أجهزة استشعار دخان الحرائق في جميع الغرف، وغيرها من الإجراءات).

بل إن الاحتمالات يمكن أن تُستخدم لرصد حالات الغش في بعض المواقف. حتى إن مؤسسة كافيون لأمن الاختبارات تتخصص فيما تصفه هي بـ«التحليل الجنائي للبيانات»،

الذي يهدف لرصد أنماط من الإجابات تنبئ بوجود حالات غش[٥]. فمثلًا تقوم تلك الشركة (والتي أنشأها أحد المطورين السابقين لاختبارات SAT) بوضع علامات استفهام على اختبارات إحدى المدارس، سواء الكتابية أو الإلكترونية، إذا ما وَجدت عددَ الإجابات الخطأ غيرَ المحتمل الحدوث بشكل يثير الريبة، عددًا تُقدَّر فرصة حدوثه بواحد في المليون. يقوم المنطق الرياضي لتلك الفكرة على حقيقة أننا لن نستطيع معرفة الكثير إذا ما أجاب رهط كبير من التلاميذ على السؤال بإجابة صحيحة. فالمفترض بالتلاميذ أن يجيبوا إجابات صحيحة، سواء كانوا يغشون أو كانوا تلاميذ نابهين. لكن عندما يجيب نفس أولئك التلاميذ إجابات خطأ، فليس من المفترض بهم أن يجيبوا نفس تلك الإجابات الخطأ بنمط متطابق. فإذا ما فعلوا ذلك فهذا يشير إلى أنهم ينسخون إجابات أحدهم الآخر (أو يتبادلون الإجابات عبر الرسائل النصية). كما تفحص الشركة تلك الاختبارات التي يبلي فيها المُختَبَر بلاءً حسنًا في الأسئلة الصعبة أكثر منه في الأسئلة السهلة (شكًّا في أنه قد حصل على الإجابات قبل الاختبار)، وفي الاختبارات التي مُسحت فيها «الإجابات الخطأ وكُتبت بعدها الإجابات الصحيحة»، بشكل أكبر من «الإجابات الصحيحة التي مُسحت وكُتبت بعدها إجابات خطأ» (لأن ذلك قد يدل على أن المعلم أو القائم على الاختبار قد عدَّل في أوراق الإجابة بعد انتهاء الاختبار.

لكنك بالطبع قادر على أن تلحظ محدودية قدرة الاحتمالات. فقد يجيب مجموعة كبيرة من المتحنين نفس الإجابات الخطأ عن طريق الصدفة، بل إننا في الواقع كلما تَفحَّصنا عددًا أكبر من المدارس وجدنا أنماط الإجابات الخطأ تلك تتأتَّى عن طريق الصدفة. فذاك الشذوذ الإحصائي لا يثبت حدوث أي فعل غير مشروع. فمثلًا حدث أن ربح ديلما كيني، ذلك الرجل ذو الخمسين ربيعًا والقاطن بولاية أطلانطا، مليون دولار في لعبة يانصيب فوري عام ٢٠٠٨، ثم عاد وربح مليونًا آخر في لعبة يانصيب فورية أخرى عام ٢٠١١[٦]. واحتمالات حدوث ذلك لذات الشخص تقترب من نسبة تبلغ ١ من ٢٥ تريليونًا. لا يمكننا أن نسجن السيد كيني بتهمة الاحتيال اعتمادًا على تلك الحسابات فقط (رغم أننا قد نبحث فيما إذا كان له أي أقارب يعملون في مسابقة يانصيب الولاية). فما الاحتمالات إلا سلاح واحد ضمن ترسانة يستلزم استخدامها حكمًا رشيدًا.

تحديد العلاقات المهمة

(أعمال التحريات الإحصائية)

هل يسبب تدخين السجائر مرض السرطان؟ إننا نملك إجابة لهذا السؤال، لكن العملية التي وصلنا بها لهذه الإجابة لم تكن حتى تقترب من البساطة والمباشرة التي تظنها. عندما نختبر فرضية علمية يقتضي المنهج العلمي حينها أن نُجري تجربة مضبوطة

يكون فيها المتغير المطلوب دراسته (كالتدخين على سبيل المثال) هو الأمر الوحيد الذي يتغير بين المجموعة التجريبية والمجموعة الضابطة. فإذا ما لاحظنا اختلافًا ملحوظًا بين مخرجات المجموعتين (كسرطان الرئة على سبيل المثال)، يمكننا حينها أن نستنتج أن ذاك المتغير هو الذي تَسبَّب في ذات المخرجات. لا يمكننا أن نجري مثل تلك التجارب على البشر. وإذا ما كانت فرضيتنا تقول بأن التدخين يسبب سرطان الرئة، فإنه لَأمرٌ غير أخلاقي أن نقسم رهطًا من المتخرجين حديثًا من الجامعة إلى مجموعتين، مدخنين وغير مدخنين، ثم ننتظر لنرى أيهم أصيب بالسرطان بعد عشرين عامًا. (يمكننا أن نجري تجربة مضبوطة على البشر إذا كانت فرضيتنا تقول بأن علاجًا معينًا أو دواءً معينًا قد يسهم في تحسين حالتهم الصحية، لكن لا يمكننا أن نعرِّض بعض الأشخاص عمدًا لتجربةٍ ما، إذا ما كنا نتوقع نتيجة سيئة).*

لعلك الآن قد وصلت إلى قناعة بأننا لا نحتاج لأن نجري مثل تلك التجربة المشكوك في أخلاقيتها لمراقبة آثار التدخين. ألَسْنا قادرين على أن نتغاضى عن ذلك المنهج العلمي المنمق، ونكتفي بالمقارنة بين معدلات الإصابة بالسرطان بين مَن شرعوا في التدخين إبان تخرُّجهم ومن لم يدخنوا، بعد مضي عشرين عامًا على تخرُّجهم من الجامعة؟

الإجابة هي لا. فمَن يدخنون ومن لا يدخنون يتباينون فيما هو أكثر من سلوكهم في ممارسة التدخين. فمثلًا يميل المدخنون لأن يتبعوا عادات سيئة أخرى، كالإسراف في معاقرة الخمر أو الأكل بشراهة، والتي من شأنها أن تسفر عن عواقب وخيمة. فإذا ما وجدنا المدخنين معتلِّي الصحة بعد عشرين عامًا، فلن نستطيع أن نجزم إذا ما كان اعتلال الصحة ذاك مردُّه إلى التدخين أو إلى تلك العادات السيئة الأخرى التي يمارسونها. كما أننا سنواجه صعوبة كبيرة في جمع البيانات التي سنبني تحليلنا عليها، لأن أصحاب الحالات المتأخرة من مرضى السرطان سيكونون أقل إقبالًا على حضور تجمعات لمِّ الشمل تلك التي سنحصل منها على بياناتهم. (كما أن مَن ماتوا بالسرطان من المدخنين لن يحضروا قطعًا). ونتيجة لهذا سيصبح أي تحليل لصحة من حضروا ذلك الاجتماع (سواء كان مرتبطًا بالتدخين أو بغيره من العادات) تحليلًا معيبًا بدرجة كبيرة؛ لأن من حضروه هم الأوفر حظًّا من الناحية الصحية. وكلما بعدت الفترة بين التخرج والتحليل، كأن يتم هذا التحليل في اجتماع لمِّ الشمل الأربعين أو الخمسين مثلًا، زادت النتائج في تطرُّفها.

لا يمكن أن نجعل من البشر فئران تجارب، ولهذا فإن الإحصائيات تبدو مشابهة كثيرًا لأعمال التحريات البوليسية. فالبيانات تُخرج لنا أدلة وأنماطًا يمكن أن تقودنا في النهاية إلى خُلاصات ذات معنًى. لعلك قد شاهدت واحدًا من تلك المسلسلات التلفزيونية الرائعة التي تحكي عن الإجراءات البوليسية على غرار سي إس آي نيويورك CSI: New York، والذي

* ذاك تبسيط مجمل لمجال أخلاقيات مهنة الطب الساحر المعقد.

نرى فيه مجموعة من المحققين وخبراء الأدلة الجنائية الوُسماء يشرعون في تأمل أدلة دقيقة – كآثار البصمة الوراثية من على عقب سيجارة أو علامات قضم على تفاحة أو إحدى ألياف دواسة أقدام في سيارة – ثم يستخدمون ذلك الدليل للإيقاع بالقاتل. إن ما يجذب المشاهدين للمسلسل هو أن هؤلاء الخبراء لا يحوزون دليلًا تقليديًّا للإيقاع بالقاتل، كشاهد عيان أو تسجيل من كاميرا مراقبة. لذلك يَلجئون إلى استدلال علمي. وهو الأمر ذاته الذي يقوم به الإحصاء. فالبيانات تزوِّدنا بأدلة غير منظمة – مثل مسرح الجريمة. والتحليل الإحصائي هو العمل التحقيقي الذي يصوغ من تلك البيانات الخام خلاصة ذات معنًى.

بعد أن تقرأ الفصل الحادي عشر، سوف تعرف أهمية المسلسل التلفزيوني الذي آمل أن أُخرجه للنور: سي إس آي تحليل الانحدار، والذي لن يختلف كثيرًا عن كل تلك المسلسلات الأخرى التي تعج بالحبكات البوليسية. وتحليل الانحدار هو أداةٌ تمكِّن الباحثين من فصل العلاقة بين اثنين من المتغيرات، كالتدخين والإصابة بالسرطان، مع التمسك بالثابت (أو ضبطه) الذي يتمثل في تأثيرات المتغيرات المهمة الأخرى، كالحمية الغذائية وممارسة الرياضة والوزن وغيرها. فعندما تقرأ في الصحيفة أنَّ تناول كعك النخالة يوميًّا يقلل من فرص الإصابة بسرطان القولون، لا تجزع ظانًّا أن الباحثين قد احتجزوا مجموعة من البشر في قبو أحد المعامل الفدرالية وأسموها المجموعة التجريبية وأجبروهم على تناول كعك النخالة، في ذات الوقت الذي يحتجزون فيه مجموعة أخرى تسمى المجموعة الضابطة في مبنًى آخر ويطعمونهم البيض ولحم الخنزير المقدد. لكن ما يحدث بالفعل هو أن الباحثين يجمعون معلومات مفصلة عن آلاف الأشخاص تتضمن معدلات تناولهم لكعك النخالة، ثم يستخدمون تحليل الانحدار في إجراءين محوريين (١) تحديد كم الارتباط الملحوظ بين تناول كعك النخالة والإصابة بسرطان القولون (كأن نجد نتيجة افتراضية مفادها أن من يأكلون كعك النخالة هم أقل إصابة بسرطان القولون بنسبة ٩ بالمائة، مع تثبيت العوامل الأخرى التي قد تؤثر في فرص الإصابة بذلك المرض)، (٢) تحديد كم أرجحية كون الارتباط بين تناول كعك النخالة وانخفاض معدل الإصابة بسرطان القولون الوارد في الدراسة محض مصادفة – مجرد شذوذ في بيانات هذه العينة من البشر – أكثر من كونه دليلًا ذا معنًى في العلاقة بين الحمية الغذائية والصحة.

لكن من سيمثلون في مسلسلي سي إس آي تحليل الانحدار، سيكونون ولا شك أكثر وسامة من أولئك الأكاديميين الذين يدرسون تلك البيانات بتمعُّن. هؤلاء الوسماء والجميلات (الذين يحوزون جميعًا درجات دكتوراه الفلسفة رغم كون أيٍّ منهم لم يتعدَّ الثالثة والعشرين من عمره) سيقومون بدراسة كميات هائلة من البيانات واستخدام أحدث الأدوات الإحصائية للإجابة عن أسئلة اجتماعية مهمة على غرار: ما أكثر أدوات مكافحة جرائم العنف فعالية؟ من الأشخاص الأعلى فرصة لأن يصيروا إرهابيين؟ سوف

نناقش في هذا الكتاب لاحقًا مفهوم النتيجة «الدالة إحصائيًّا»، وهو ما يعني أن التحليل قد كشف عن أن الارتباط بين متغيرين لا يميل لأن يكون محض مصادفة فقط. وهذا النوع من النتائج الإحصائية يَعُدُّه الأكاديميون والباحثون «أمرًا مثيرًا للريبة». في مسلسلي أتخيل باحثة أكاديمية تعمل في مختبر الحاسوب في وقت متأخر من الليل، لأنها مرتبطة وقت النهار بالتزامات عضويتها في منتخب الولايات المتحدة لكرة الطائرة الشاطئية. ثم تطبع تحليلها الإحصائي فتجد فيه ما كانت تنشده بالضبط؛ ألا وهو علاقة دالة إحصائيًّا قائمة داخل البيانات بين متغير ما، افترضت هي أهميته، وبين إصابة البشر بالتوحد. هنا يتعين عليها أن تطلع الآخرين على هذا الكشف الهائل فورًا.

تهرع الباحثة حاملة نتائج تحليلها المطبوعة تشق أروقة المبنى، يبطئها نوعًا ما حذاؤها ذو الكعب العالي وتنورتها السوداء القصيرة بعض الشيء، حتى تجد زميلها الذي لا نجد تفسيرًا للياقته البدنية وسمرة بشرته المكتسبة، رغم أنه يعمل في قبو مختبر الكمبيوتر لأربع عشرة ساعة يوميًّا، فتعرض عليه النتائج. يعبث هو بأصابعه في لحيته المربعة المشذبة بعناية، ثم يفتح درج مكتبه فيستلُّ منه مسدسه الجلوك عيار التسعة ميلليمترات ويدسه في غمد كتفه أسفل سترة حلته التي تحمل علامة هيوجو بوس، والتي يبلغ ثمنها خمسة آلاف دولار (وهو أمر آخر لا نجد له تفسيرًا بما أن راتبه كأكاديمي مبتدئ يقدر بــ ٣٨٠٠٠ دولار أمريكي سنويًّا). ثم يمضي خبيرَا تحليل الانحدار معًا بخفة قاصدَين مكتب رئيسهما، الذي هو رجل أشيب من قدامى المحاربين نجح في التعافي من علاقات فاشلة ومشكلات تتعلق بإدمان الكحوليات

حسنًا، لست مضطرًّا لأن تعيش تلك الدراما التلفزيونية كي تدرك أهمية هذا النوع من البحث الإحصائي. فكل القضايا الاجتماعية التي تشغلنا تقريبًا قد استقينا المعلومات عنها من ذلك التحليل النظامي لمجموعات كبيرة من البيانات (في العديد من الحالات يلعب جمع البيانات المتعلقة بالقضية، والذي هو إجراء مكلف ماديًّا ومستهلك للوقت، دورًا أساسيًّا في العملية كما سنبين في الفصل السابع). قد أكون بالغت في تزيين صورة شخصيات مسلسلي ذلك، لكنني لم أبالغ في وصف الأسئلة المهمة التي يتعين عليهم دراستها. فهناك من المؤلفات الأكاديمية ما يدرس الإرهابيين والانتحاريين من مفجري القنابل، وهو موضوع تتعذر دراسته على عينات من البشر (أو حتى على فئران التجارب). واحد من تلك الكتب مُعنون What Makes a Terrorist، كتبه واحد من أساتذتي الذين درَّسوني الإحصاء في مرحلة الدراسات العليا. يستخلص الكتاب نتائجه من بيانات جُمعت عن الهجمات الإرهابية التي حدثت حول العالم. وقد جاءت نتائج العينة المدروسة لتقول بأن الإرهابيين ليسوا فقراء مُعدمين ولا جهلة لم يتلقوا تعليمًا جيدًا. بل استخلص المؤلف، آلان كروجر أستاذ الاقتصاد بجامعة برنستون، أن: «الإرهابيين عادة ما ينحدرون من عائلات

حظيت بتعليم رفيع المستوى، سواء كانت من الطبقات الوسطى أو من العائلات ذات الدخل المادي العالي»[٧].

لماذا؟ يكشف هذا السؤال عن أحد النقائص التي تعتري تحليل الانحدار. فإننا قادرون على أن نحدد الارتباط القوي القائم بين متغيرَيْنِ مستخدِمِينَ التحليل الإحصائي، لكننا لسنا قادرين بالضرورة على تفسير وجود ذلك الارتباط من الأساس، بل وفي بعض الحالات لا نستطيع أن نتيقن من كون تلك العلاقة علاقة سببية؛ بمعنى أن أي تغيُّر يطرأ على أيٍّ من المتغيرين يسبب بالفعل تغيرًا في الآخر. ففي حالة الإرهاب يفترض البروفيسور كروجر أنه بما أن الإرهابيين تُحركهم أهداف سياسية، فإن أولئك الأعلى تعليمًا والأكثر ثراءً هم الذين يملكون الحافز الأكبر لتغيير المجتمع. كما أن هؤلاء الأشخاص قد تكون صدورهم موغَرة من جراء قمع حريتهم، وهو عامل آخر يرتبط بالإرهاب. فقد أورد كروجر في دراسته أن البلدان ذات المستويات العالية من القمع السياسي لها نصيب أكبر من الأنشطة الإرهابية (مع تثبيت العوامل الأخرى).

تقودني تلك المناقشة مرة أخرى إلى السؤال الذي هو عنوان الفصل: ما المقصد؟ ليس المقصد في الحسابات أو في إبهار الزملاء والأصدقاء بالأساليب الإحصائية المتطورة، بل المقصد هو تعلُّم تلك الأمور التي تُثري حياتنا بالمعلومات.

أكاذيب، أكاذيب لعينة وإحصائيات

نادرًا ما يكشف التحليل الإحصائي عن «الحقيقة»، حتى عندما تتهيأ أفضل الظروف. فإننا عادةً ما نبني قضية ظرفية قائمة على بيانات غير مثالية. ونتيجة لهذا يجد أصحاب الأمانة العلمية لديهم العديد من الأسباب التي تجعلهم يختلفون حول النتائج الإحصائية أو ما تنطوي عليه من مضامين. فقد نختلف على السؤال المطروح للإجابة عن المستوى الأكثر أساسية. فقد يظل مشجعو الألعاب الرياضية يتجادلون إلى الأبد حول «أفضل لاعب بيسبول جاد به الزمان على الإطلاق»، لأنه ليس هناك تعريف موضوعي لمفهوم «الأفضل». والإحصائيات الوصفية المنمقة يمكن أن تأتي بمعلومات حول هذا السؤال، لكنها لن تجيب عنه إجابة قاطعة. وكما سنرى في الفصل التالي، تتعرض الكثير من القضايا الاجتماعية المهمة لنفس هذه المشكلة. ما الذي يحدث لصحة الحالة الاقتصادية للطبقة الوسطى الأمريكية؟ تعتمد إجابة هذا السؤال على تعريف كلٍّ من مصطلحي «الطبقة الوسطى» و«صحة الحالة الاقتصادية».

هناك حدود تحدُّ عملية جمع البيانات، وكذلك تحدُّ التجارب التي يمكن أن نجريها. فدراسة آلان كروجر للإرهابيين لم تتتبَّع الآلاف من الشباب على مدى عشرات السنين لملاحظة أيٍّ منهم تحوَّل إلى إرهابي. فهذا ليس ممكنًا. وكذلك لا نستطيع أن نخلق دولتين

متطابقتين في كل شيء سوى أن إحداهما دولة قمعية والأخرى ليست كذلك، ثم نقارن بين عدد مفجِّري القنابل من الانتحاريين الذين يبرزون من كل واحدة منهما. وحتى عندما نجري تجارب مضبوطة واسعة النطاق على البشر، نجدها ليست بالهينة ولا الزهيدة التكلفة. كان الباحثون قد قاموا بدراسة متسعة النطاق تبحث فيما إذا كانت الصلاة تحدُّ من مضاعفات ما بعد العمليات الجراحية، وهو السؤال الذي أثرناه ضمن أسئلة أخرى في بداية هذا الفصل. بلغت تكلفة تلك الدراسة ٢٫٤ مليون دولار أمريكي (أما عن النتائج فعليك أن تنتظرها في الفصل الثالث عشر).

لوزير الدفاع الأمريكي دونالد رمسفيلد عبارة شهيرة تقول: «إنك تذهب للحرب بالجيش الذي تملكه – لا الجيش الذي تريده أو تتمنى أن تحوزه يومًا». ومهما كان رأيك في رمسفيلد (وفي حرب العراق التي كان يصفها بقوله هذا)، فإن هذا القول المأثور ينطبق على البحث كذلك. فإننا نجري البحث الإحصائي مستخدمين أفضل البيانات والمنهجيات والموارد المتاحة. لا يشبه هذا التحليل عملية جمعٍ حسابيٍّ أو قسمة مطولة يخرج فيها الأسلوب الصحيح بالإجابة «الصحيحة» ويكون فيه الحاسوب أكثر دقة وأقل عرضة للخطأ من الإنسان، بل إن التحليل الإحصائي أشبه بأعمال تحريات بوليسية بارعة (وهو ما يبشر بنجاح تجاري لمسلسل سي إس آي تحليل الانحدار). لكن أذكياء الناس والأمناء منهم سوف يظلون يختلفون حول ما تحاول البيانات أن تخبرنا به.

لكن من قال إن كلَّ مَن يستخدم الإحصاء ذكي وأمين؟ كما ذكرنا من قبل جاء هذا الكتاب امتدادًا لكتاب How to Lie With Statistics، وترجمة عنوانه حرفيًّا هو «كيف تكذب بالإحصاء»، وهو كتاب صدرت طبعته الأولى عام ١٩٥٤ وبيعت منه أكثر من مليون نسخة. فالحقيقة أنك تستطيع أن تكذب بالإحصاء. أو قد تقع في أخطاء غير مقصودة. وفي كلتا الحالتين يتلبس الدقةَ الحسابيةَ المتصلةَ بالتحليل الإحصائي بعضُ الهراء الموغل في التفاهة. سوف يتناول هذا الكتاب العديد من أكثر الأخطاء والتزييفات الإحصائية شيوعًا (حتى يتسنى لك التعرف عليها لا استخدامها).

وهكذا كي نعود لعنوان الفصل: ما المقصد من تعلم الإحصاء؟

المقصد هو تلخيص كميات هائلة من البيانات.

المقصد هو اتخاذ قرارات أفضل.

المقصد هو الإجابة عن الأسئلة الاجتماعية المهمة.

المقصد هو تحديد الأنماط التي تُعلي من جودة طرائقنا التي نتصرف بها في كل أمورنا؛ من بيع حفاضات الأطفال حتى اعتقال المجرمين.

المقصد هو الإيقاع بالغشاشين ومحاكمة المجرمين.

المقصد هو تقييم فعالية السياسات والبرامج والعقاقير الدوائية والإجراءات الطبية وغيرها من الجهود الإبداعية.

المقصد هو تعرف أولئك الأوغاد الذين يستخدمون نفس تلك الأداة الفعالة لغاياتهم المشينة.

إذا ما كنت ترى في نفسك القدرة على القيام بكل ما سبق وأنت ترتدي حلة من ماركة هيوجو بوس، أو ترتدين تنورة سوداء قصيرة، فربما كنتَ أو كنتِ النجم القادم لمسلسل سي إس آي تحليل الانحدار.

الفصل الثاني

الإحصائيات الوصفية

من أفضل لاعبي البيسبول على مر التاريخ؟

لنتأمل للحظة اثنين من الأسئلة اللذين لا تتبدى بينهما أية علاقة: (١) ما الذي يحدث للصحة الاقتصادية للطبقة الوسطى الأمريكية؟ و(٢) من أفضل لاعبي البيسبول على مر التاريخ؟

أول هذين السؤالين هو سؤال مهم للغاية، فهو عادة ما يكون في القلب من الحملات الدعائية للانتخابات الرئاسية وغيرها من الحركات الاجتماعية. فالطبقة الوسطى هي قلب أمريكا؛ لذا فإن صحة هذه المجموعة تعد مؤشرًا مهمًّا على الصحة الاقتصادية للأمة بأسرها. أما السؤال الثاني فهو سؤال تافه (بالمعنى الحرفي للكلمة)، غير أن مشجعي البيسبول المتحمسين قد ينخرطون في الجدال حوله لأبد الدهر. لكن ما يشترك فيه السؤالان هو أن كليهما يمكن أن يُستخدم لتبيان مواطن القوة ومواطن العوار في الإحصائيات الوصفية، والتي هي عبارة عن أرقام وحسابات نستخدمها لتلخيص البيانات الخام.

إذا ما أردتُ أن أبيِّن كيف أن ديريك جيتر لاعب بيسبول عظيم، فإنني قادر على أن أشرع في وصف كل ضربة ضربها بمضربه في كل مباراة في بطولة كبرى. لكن هذه الأوصاف بيانات خام، وسوف تستلزم منا وقتًا لهضمها؛ حيث إن جيتر لعب لسبعة عشر موسمًا مع فريق النيويورك يانكيز، وضرب ٩٨٦٨ ضربة بمضربه.

أو يمكنني أن أقول لكم إنه بنهاية موسم عام ٢٠١١ كان متوسط ضربات الكرات التي ضربها ديريك جيتر خلال مسيرته الرياضية ٣١٣ ضربة، هذه إحصائية وصفية أو «إحصائية تلخيصية».

متوسط ضرب الكرات هذا هو عبارة عن تبسيط إجمالي للسبعة عشر موسمًا التي لعبها جيتر. هو أمر سهل الفهم أنيق في بساطته ومحدود فيما يفصح لنا عنه. لكن الخبراء بلعبة البيسبول لديهم حزمة من الإحصائيات الوصفية التي يَعدُّونها أكثر قيمة من متوسط

ضرب الكرات. لقد هاتفتُ ستيف موير رئيس مؤسسة بيسبول إنفو سوليوشنز (وهي مؤسسة تقوم بتقديم بيانات خام للكتب التي تتناول الإحصائيات المتعلقة بلعبة البيسبول على غرار كتاب Moneyball) كي أسأله السؤالين التاليين: (١) ما أكثر الإحصائيات أهمية في تقييم الموهبة في لعبة البيسبول؟ (٢) من أفضل لاعبي البيسبول على مر التاريخ؟ ولسوف أطلعك على إجاباته عندما يحين السياق المناسب لذلك.

لكن لنَعُدِ الآن لذلك الموضوع الذي يقل تفاهة عن سابقه؛ ألا وهو الصحة الاقتصادية للطبقة الوسطى. لو طلبنا المثالية لبحثنا عن مكافئ لمتوسط ضرب الكرات أو ربما شيء أفضل منه، فما نحتاجه هو مقياس بسيط ولكن دقيق نقيس به كيف تتغير الرفاهة الاقتصادية للعامل الأمريكي المنتمي إلى الطبقة الوسطى خلال الزمن الراهن. أن نحدد هل يزداد الناس الذين نراهم ينتمون إلى الطبقة الوسطى ثراءً أو يزدادون فقرًا أو أنهم كمن يجري في مكانه؟ قد تكون الطريقة المعقولة – والتي هي ليست بالطريقة «الصحيحة» أبدًا – هي حساب التغير في نصيب الفرد من الدخل في الولايات المتحدة الأمريكية على مدى جيل كامل، أي نحو الثلاثين عامًا. ونصيب الفرد من الدخل هو متوسط بسيط ينتج عن عملية قسمة إجمالي الدخل على عدد السكان. وبهذا المقياس نجد متوسط الدخل في الولايات المتحدة قد صعد من ٧٧٨٧ دولارًا عام ١٩٨٠ إلى ٢٦٤٨٧ دولارًا عام ٢٠١٠ (وهو آخر عام جمعت فيه الحكومة البيانات)[١]. هنيئًا لنا قد حققنا الهدف.

لكن هناك مشكلة، وهي أن حساباتي السريعة برغم صحتها من الناحية الفنية فإنها خطأ تمامًا إذا ما نظرنا لها كإجابة للسؤال الذي طرحته في البداية. فهي، أولًا، لم تراعِ معدلات التضخم في الأرقام السابقة. (فنصيب الفرد من الدخل المقدر بـ ٧٧٨٧ دولارًا عام ١٩٨٠ يعادل نحو ١٩٦٠٠ دولار طبقًا لقيمة الدولار عام ٢٠١٠). هذا حل سريع لتلك المشكلة، لكن المشكلة الأكبر تكمن في كون متوسط الدخل في أمريكا لا يساوي دخل الأمريكي العادي. دعنا نحلل تلك العبارة القصيرة الحاذقة.

إن إحصائية نصيب الفرد في الدخل لا تفعل أكثر من أن تقسم إجمالي الدخل الذي جنته الدولة على عدد سكانها، وهو الأمر الذي لا يمنحنا أي معلومة عما يربحه كل فرد تحديدًا من هذا الدخل، سواء عام ١٩٨٠ أو عام ٢٠١٠. فكما يقول أولئك الرفاق في حركة احتلوا وول ستريت، إن النمو الهائل في دخل الواحد بالمائة الذين يعتلون هرم الدخل يمكن أن يزيد إحصائية نصيب الفرد من الدخل بصورة كبيرة، لكن دون أن يمنح التسعة والتسعين بالمائة الآخرين أية نقود. بعبارة أخرى نقول إن متوسط الدخل قد يرتفع لكن دون أن يساعد الأمريكي العادي.

وكما فعلت مع إجابة السؤال المتعلق بإحصائية البيسبول، فقد نشدت إجابة للسؤال الخاص بالصحة الاقتصادية للطبقة المتوسطة الأمريكية من مصادر خارجية خبيرة.

فسألت اثنين من كبار المتخصصين في الاقتصاد العمالي، كان من بينهما كبير مستشاري الرئيس أوباما الاقتصاديين، عن الإحصائيات الوصفية التي يستخدمونها عندما يريدون تقييم الرفاهة الاقتصادية للأمريكي العادي. ولسوف أطلعكم على إجابتهما أيضًا بعد أن أنتهي من اصطحابكم في جولة سريعة لاستكشاف الإحصائيات الوصفية؛ لأنها حينها ستصير ذات معنًى أكبر لديكم.

عندما يتعامل المرء مع البيانات، سواء كان ذلك في لعبة البيسبول أو في نصيب المرء من الدخل، تصير مهمته الأساسية هي تلخيص قدر هائل من المعلومات. هناك نحو ٣٣٠ مليون نسمة يقطنون الولايات المتحدة. ويمكن أن تتوافر لدينا كل المعلومات التي نحتاجها عن الصحة الاقتصادية للدولة إذا ما توافر لدينا جدول لكل مواطن أمريكي يحمل اسمه وسجل دخله على مر الأعوام، لكن ذلك الجدول في ذات الوقت سيفتقر إلى العملية؛ بحيث لن يخبرنا بأي شيء على الإطلاق. تكمن المفارقة هنا في كون كثرة البيانات تمثل عادة وضوحًا أقل؛ لذا وجب علينا التبسيط، وأن نجري عمليات حسابية قادرة على اختزال فيض معقد من البيانات إلى حفنة من الأرقام التي تصف تلك البيانات، تمامًا كما نلخص أداءً أولمبيًّا معقدًا متعدد الأوجه في رقم واحد على غرار ٩٫٨.

الخبر الطيب هنا هو أن هذه الإحصائيات الوصفية تمنحنا تلخيصًا سهل الاستيعاب ذا معنًى للظواهر التي ندرسها. وهذه هي الفكرة التي يتناولها هذا الفصل. أما الخبر السيء فهو أن أي تبسيط يتيح فرصًا لإساءة الاستخدام. فالإحصائيات الوصفية يمكن أن تصير – كحساب في موقع مواعدة إلكتروني – دقيقة من الناحية الحرفية، لكنها مع هذا مضللة جدًّا.

لنفترض أنك جالس في محل عملك تتصفح شبكة الإنترنت هائمًا بلا هدف، فتعثر في سجل يومي جاذب للاهتمام لزيجة كيم كارداشيان الفاشلة بلاعب كرة السلة المحترف كريس همفريز، والتي استمرت لاثنين وسبعين يومًا. وبعد أن انتهيت من قراءة اليوم السابع من سجل هذه الزيجة، يهل عليك مديرك حاملًا ملفين عملاقين مكتظين بالبيانات. واحد من هذين الملفين يحتوي على معلومات طلبات استخدام حق الضمان لكل واحدة من الـ ٥٧٣٣٤ طابعة ليزر التي باعتها مؤسستك العام الماضي. (في ذلك الملف توثيق لعدد المشكلات المتعلقة بجودة الطابعة، والتي تم الإبلاغ عنها خلال فترة الضمان). أما الملف الثاني فيحوي نفس نوع المعلومات، لكنها تخص الـ ٩٩٤٧٧٣ طابعة ليزر التي باعها أشد منافسيكم خلال نفس الفترة الزمنية. يريد مديرك أن يعرف كيف هي طابعات المؤسسة بالمقارنة مع منافسكم من حيث الجودة.

لحسن الحظ يتوافر في الحاسوب، ذلك الذي كنت تشاهد عليه قصة زيجة كارداشيان، حزمة من برامج الإحصائيات الأساسية، لكن من أين تبدأ؟ سيدفعك

إحساسك الغريزي إلى البداية الصحيحة على الأرجح: فأولى خطوات الوصف تتمثل عادة في إيجاد مقياس للمجموعة «الوسطى» من البيانات، أو التي يصفها المختصون بالإحصاء بـ «النزعة المركزية». ما التجربة المعتادة التي يختبرها مستخدمو طابعاتكم مقارنة بمن يجربون طابعات منافسكم من حيث الجودة؟ يعتبر أكثر المقاييس أساسية لقياس «أوسط» ذلك التوزيع هو المتوسط. في هذه الحالة نريد أن نعرف متوسط عدد المشكلات المتعلقة بالجداول لكل طابعة باعتها مؤسستك، وكل طابعة باعتها المؤسسة التي تنافسكم. تستطيع بكل بساطة أن تحصي عدد المشكلات المتعلقة بالجودة التي تم الإبلاغ عنها خلال فترة ضمان الصيانة، ثم تقسمها على عدد الطابعات المَبيعة. (تَذكَّرْ أن طابعة واحدة يمكن أن تعاني من عدة مشكلات خلال فترة ضمان الصيانة). سوف تفعل ذلك مع كل مؤسسة، خالقًا إحصائية وصفية مهمة، ألا وهي متوسط عدد مشكلات الجودة في كل طابعة مَبيعة.

لنفترض أنك وجدت طابعات منافسكم ذات متوسط يقدر بـ ٢,٨ للمشكلات المتعلقة بالجودة لكل طابعاتها خلال فترة ضمان الصيانة، مقارنة بـ ٩,١ الذي هو متوسط عيوب طابعاتكم التي أبلغ عنها. كان هذا سهلًا، فها أنت قد أخذت معلومات عن مليون طابعة باعتها مؤسستين مختلفتين وقطَّرتها حتى وصلت إلى لب المشكلة؛ ألا وهي أن طابعاتكم كثيرة الأعطال. ومن الواضح الآن أن عليك أن تبعث برسالة إلكترونية قصيرة لمديرك تقدِّر فيها حجم تلك الفجوة الحادثة في الجودة، وبعدها تعود لمتابعة سجل اليوم الثامن من زيجة كيم كارداشيان.

أو ربما لن تستطيع ذلك. لقد تعمدت الغموض سابقًا عندما أشرت إلى «وسط» التوزيع. يتضح أن المتوسط ينطوي على بعض المشكلات في هذا الصدد، منها أنه عرضة لبعض تشوهات تُسببها «القيم المتطرفة»، والتي هي مشاهدات تتموضع بعيدًا عن المركز. لتقريب هذا المفهوم إلى ذهنك تخيل عشرة أشخاص يجلسون على مقاعد دوارة في حانة يرتادها أفراد الطبقة المتوسطة بمدينة سياتل، كل واحد من أولئك العشرة يتقاضى ٣٥٠٠٠ دولار سنويًّا، وهو ما يجعل متوسط الدخل السنوي للمجموعة ٣٥٠٠٠ دولار. تخيل أن بيل جيتس يدلف إلى ذات الحانة وعلى كتفه ببغاء ناطقة. (لا فائدة من الببغاء في هذا المثال في الواقع سوى جعله أكثر إثارة). دعنا نفترض لتحقيق غرض هذا المثال أن الدخل السنوي لبيل جيتس يقدر بمليار دولار. عندما يجلس بيل على المقعد الحادي عشر سوف يرتفع متوسط الدخل السنوي لزبائن الحانة إلى نحو ٩١ مليون دولار، لكن كما هو واضح، لم يزدد أيٌّ من الزبائن العشرة الأوائل ثراءً بأي حال (رغم أن بيل قد يكون قد ابتاع للجميع الشراب مرة أو مرتين). فإذا ما وصفنا زبائن الحانة بكونهم ذوي متوسط دخل سنوي يبلغ ٩١ مليون دولار، فإن هذه المقولة صحيحة إحصائيًّا لكنها فادحة

في تضليلها. لكن ذلك المكان ليس من النوعية التي يتجمع فيها أصحاب الثروات التي تقدر بالملايين، وإنما هي حانة يتجمع فيها مجموعة من الرجال ذوي الدخول المنخفضة نسبيًّا، والذين تَصادَف جلوسهم بجوار بيل جيتس وببغائه الناطقة. وحساسية المتوسط هذه للقيم المتطرفة هي السبب الذي لأجله لا ينبغي لنا أن نقيس مدى الصحة الاقتصادية للطبقة الوسطى الأمريكية من خلال النظر إلى نصيب الفرد من الدخل السنوي. ولأنه كان هناك انفجار في نمو الدخول عند النهاية العليا للتوزيع – كما في دخول المديرين التنفيذيين ومديري المحافظ الوقائية والرياضيين من أمثال ديريك جيتر – فقد يحدث انحراف شديد في متوسط الدخل في الولايات المتحدة، مما يجعله قريب الشبه بحالة كراسي الحانة الدوارة التي يجلس بيل جيتس على آخر واحد منها.

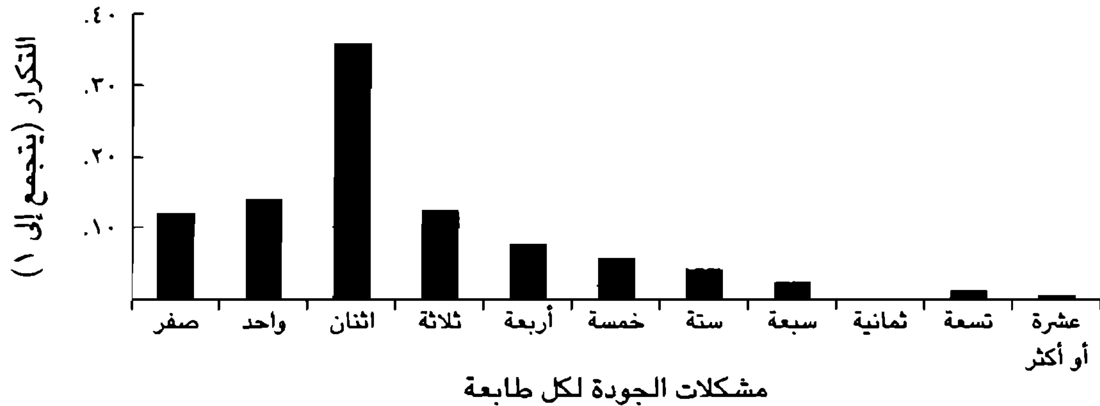
ولهذا فإن لدينا إحصائية أخرى تشير هي أيضًا إلى «وسط» التوزيع ولكن بشكل آخر، ألا وهي «الوسيط». والوسيط هو النقطة التي تقسم التوزيع إلى قسمين، مما يعني أن نصف المشاهدات يقع فوق الوسيط، ويقع النصف الآخر أسفله. (وإذا جاء عدد المشاهدات زوجيًّا، يكون الوسيط هو النقطة الوسطى بين مشاهدتين من المنتصف). إذا رجعنا إلى مثال مقاعد الحانة الدوارة، فسيكون الدخل الوسيط للعشرة أشخاص الذين جلسوا على المقاعد منذ البداية ٣٥٠٠٠ دولار، وعندما يلتحق بهم بيل جيتس وببغاؤه سيظل الرقم السابق على حاله دون تغير. وإذا ما تخيلت زبائن الحانة يصطفون على مقاعدهم بترتيب تصاعدي طبقًا لدخولهم، فسوف يمثل الزبون الجالس على المقعد السادس وسيط دخل المجموعة. وإذا ما جاء وارن بافت وجلس على المقعد الثاني عشر بجوار بيل جيتس فسيظل الوسيط كما هو*.

في حالة التوزيعات التي تخلو من قيم بالغة التطرف يتساوى المتوسط والوسيط. لقد أوردتُ تلخيصًا افتراضيًّا لبيانات الجودة الخاصة بطابعات المؤسسة المنافسة. بتعبير أدق لقد وضعت البيانات فيما يعرف بالتوزيع التكراري. فيه نجد عدد مشكلات الجودة لكل طابعة موزعة بطول أسفل الشكل، ويمثل ارتفاع كل عمود نسبة الطابعات المَبيعة مع هذا العدد من مشكلات الجودة. فمثلًا حدث فيما نسبته ٣٦ بالمائة من طابعات المؤسسة المنافسة عُطلان من أعطال الجودة خلال فترة ضمان الصيانة. ولأن التوزيع يشمل جميع المخرجات الممكنة المتعلقة بالجودة، ومن بينها اللاأعطال، فإن تلك النسب لا بد أن تتجمع لتصير ١ (أو مائة بالمائة).

---

* إذا شغل المقعد الثاني عشر سوف يصير الوسيط النقطة الوسط بين دخل الرجل الجالس على المقعد السادس ودخل الجالس على المقعد السابع. وبما أن كليهما يتقاضيان ٣٥٠٠٠ دولار سنويًّا يكون الوسيط هنا هو ٣٥٠٠٠ دولار. أما إذا كان أحدهما يتقاضى ٣٥٠٠٠ والآخر يتقاضى ٣٦٠٠٠، فإن وسيط المجموعة كاملة سيكون ٣٥٥٠٠.

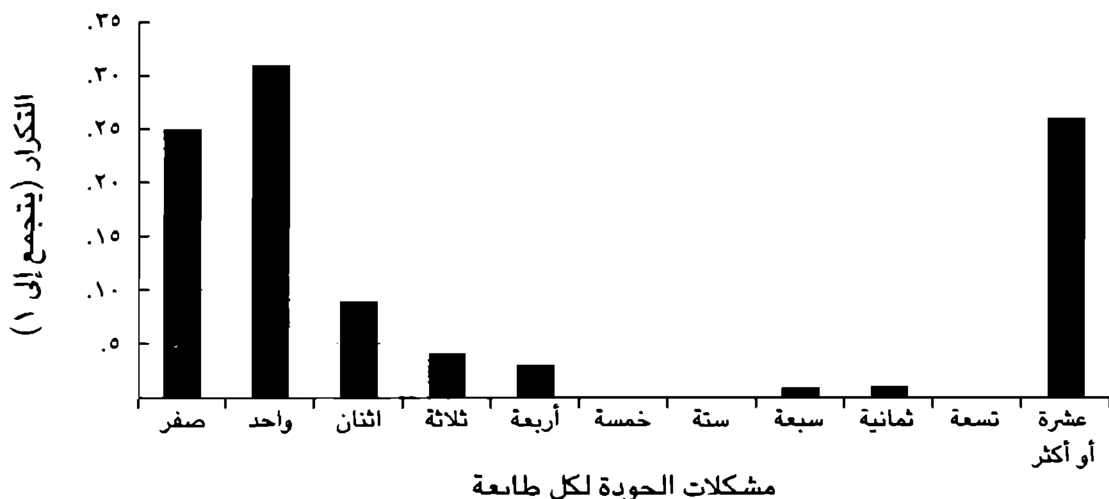
التوزيع التكراري لشكاوى المشكلات المتعلقة بالجودة في طابعات المؤسسة المنافسة



ولأن التوزيع يكاد يكون متناسقًا، فقد صار الوسط والوسيط قريبين نوعًا ما أحدهما من الآخر. وكون التوزيع ينحرف قليلًا تجاه اليمين سببه ذلك العدد البسيط من الطابعات التي سُجل بها الكثير من الأعطال المتعلقة بالجودة. هذه القيم المتطرفة تحرِّك المتوسط حركة بسيطة تجاه اليمين، لكنها غير ذات تأثير على الوسيط. لنفترض أنك، وقبل أن تهرع بتقرير الجودة إلى رئيسك في العمل، قررت أن تحسب رقم الوسيط للمشكلات المتعلقة بالجودة لطابعات مؤسستك والمؤسسة المنافسة. سوف تحصل على النتيجة بمجرد ضغطات قليلة على مفاتيح الحاسوب. سيبلغ الرقم الوسيط لشكاوى الجودة الخاصة بطابعات المنافس ٢، والرقم الوسيط لشكاوى الجودة الخاصة بطابعاتك ١

ماذا؟ رقم الوسيط للشكاوى الخاصة بكل طابعة من طابعاتك أقل من منافسك. ولأن سجل زيجة كارداشيان غدَا مملًّا، ولأن تلك النتيجة قد أثارت فضولك، فسوف تقوم بطباعة توزيع تكراري لمشكلات الجودة الخاصة بك.

التوزيع التكراري للشكاوى المتعلقة بالجودة الخاصة بمؤسستك



ما يتضح هنا هو أن شركتك لا تعاني من مشكلة عامة تتعلق بجودة طابعاتها، وإنما تعاني مما يعرف بمشكلة «ليمونية»*؛ حيث تَعرَّض عدد قليل من الطابعات إلى كم كبير من المشكلات المتعلقة بالجودة. وهذه القيم المتطرفة تقوم بتضخيم المتوسط، لكنها لا تغيِّر في الوسيط. وأهم من ذلك من وجهة نظرٍ إنتاجية أنك لا تحتاج لإعادة تهيئة العملية التصنيعية بأسرها، وإنما كل ما عليك فعله هو تحديد المكان الذي تأتي منه تلك الطابعات المنخفضة الجودة وإصلاح ذلك الأمر.**

إن المتوسط والوسيط كليهما ليسا صعبَي الحساب، لكن الفكرة هنا هي تحديد أيُّ مقاييس «المنتصف» أكثر دقة في موقف بعينه (وهي ظاهرة يسهل استغلالها). لكن في ذات الوقت هناك للوسيط بعض الأقرباء النافعين. فكما قلنا سابقًا يَقسم الوسيطُ التوزيعَ لنصفين. لكن التوزيع يمكن أن تزداد قسمته إلى أرباع أو رُبَيْعات. يتكون الرُّبَيع الأول من نسبة الـ ٢٥ الدنيا من المشاهدات، أما الرُّبَيع الثاني فيتكون من نسبة الـ ٢٥ التي تليها، وهلم جرًّا. أو قد يقسم التوزيع إلى شرائح عشرية يمثل كلٌّ منها ١٠ بالمائة من المشاهدات. (إذا كان دخلك في الشريحة العشرية العليا من توزيع الدخل الأمريكي، فستجد نفسك تجني أكثر من ٩٠ بالمائة من رفاقك العاملين). بل إننا قد نتمادى فنقسم التوزيع إلى فئات مئوية أو نسب مئوية. حيث تمثل كل نسبة مئوية واحدًا بالمائة من التوزيع، وتمثل النسبة المئوية الأولى قاع التوزيع، بينما تمثل النسبة المئوية رقم ٩٩ قمته.

إن الفائدة من مثل هذا النوع من الإحصائيات الوصفية هي أنها تصف وضع مشاهدة بعينها مقارنةً بوضع غيرها. فإذا ما قلت لك إن نتائج ابنك في اختبار القراءة والفهم قد حلت في النسبة الثالثة، فسوف تدرك فورًا أن على الأسرة أن تمضي المزيد من الوقت في المكتبة. لن يتعين عليك حينها أن تعرف أي شيء عن الاختبار نفسه أو عن عدد الأسئلة التي أجاب عنها ابنك بإجابات صحيحة. فالنسبة المئوية تمنحك تراتبية لما أحرزه ابنك مقارنة ببقية رفاقه ممن خضعوا للاختبار. وإذا ما كان الاختبار سهلًا، فسيزيد عدد الإجابات الصحيحة التي أحرزها الطلبة، لكن إجابات ابنك الصحيحة ستكون أقل من مثيلتها لدى أقرانه. أما إذا كان الاختبار بالغ الصعوبة، فستكون إجابات الطلبة الصحيحة قليلة جدًّا ولكن ستظل إجابات ابنك أقلهم.

هنا فرصة مناسبة لعرض بعض المصطلحات المفيدة. إن النتيجة أو الرقم أو العدد «المطلق» يحمل قيمة نابعة من داخله. إذا ما ضربتُ كرة الجولف ٨٣ مرة لأُدخل الكرة

* الشكلة الليمونية هي مشكلة تنشأ في الإنتاج أو الاستثمار نتيجة لاضطراب قائم على عدم تطابق معلوماتي (المترجم).

** معلومة جديدة متعلقة بالتصنيع: اتضح أن جميع الطابعات المعطوبة تقريبًا صُنعت في مصنع بكنتاكي قام العاملون فيه بتفكيك أجزاء من خط التصنيع ليبنوا بها ماكينة لتقطير خمر البوربون. وقد كان العمال المخمورون دائمًا مع الأجزاء الناقصة من خط التصنيع هما السببان اللذان تضافرا لينتجا تلك الطابعات الفقيرة الجودة.

في ثماني عشرة حفرة، فهذا عدد مطلق. وقد يكون هذا في يوم تبلغ درجة الحرارة فيه ٥٨ درجة، والذي هو أيضًا عدد مطلق. فالأعداد المطلقة يمكن أن تفسَّر خارج أي سياق ودون معلومات إضافية. فعندما أقول لك إنني ضربت الكرة ٨٣ مرة، فلا تحتاج أنت أن تعرف عدد ضربات اللاعبين الآخرين كي تقيم أدائي. (قد يكون في ذلك الأمر استثناء إذا ما كانت الظروف الجوية سيئة للغاية، أو إذا كان ملعب الجولف وعرًا جدًّا أو سهلًا جدًّا). لكن إذا ما حللت في المركز التاسع في بطولة الجولف فذاك يعتبر إحصائية نسبية. والقيمة «النسبية» أو العدد النسبي ليس له من معنًى إلا بالمقارنة مع شيء آخر أو في ضوء سياق أوسع، كأنْ أقارَن بثمانية من لاعبي الجولف يضربون الكرات أفضل مني. معظم الاختبارات الموحدة تخرج بنتائج لا تحمل من مدلولات إلا كإحصائيات نسبية. فإذا ما قلت لك إن تلميذًا في الصف الثالث بمدرسة إلينوي الابتدائية أحرز ٤٣ نقطة من ٦٠ في حصة الرياضيات باختبار ولاية إيلينوي، فإن هذا العدد المطلق لا يحمل مدلولًا في حد ذاته. لكنني عندما أحوِّله إلى نسبة مئوية – بمعنى أن أضع تلك النتيجة الخام في توزيع مع نتائج اختبار الرياضيات لجميع تلاميذ الصف الثالث بولاية إيلينوي – سيكون له حينها مدلول كبير. وإذا ما كانت الإجابات الـ ٤٣ الصحيحة تدخل في نطاق النسبة المئوية الـ ٨٣، فمعنى هذا أن هذا التلميذ يفوق معظم أقرانه في جميع أرجاء الولاية. أما إذا كان ضمن نطاق النسبة المئوية الثامنة، فهذا يعني أنه متعثر في الدراسة. وفي تلك الحالة تكون النسبة المئوية (النتيجة النسبية) أكثر دلالة من عدد الإجابات الصحيحة (النتيجة المطلقة).

واحدة من الإحصائيات الأخرى التي قد تساعدنا على وصف فوضى الأرقام تلك، فتجعلها مرتَّبةً سهلةَ الفهم، هي الانحراف المعياري، والذي هو مقياس لمدى تشتت البيانات عن متوسطها. بعبارة أخرى نقول إنه يقيس مدى تشتت تلك المشاهدات. لنفترض أنني جمعت بيانات عن أوزان ٢٥٠ راكبًا مسافرين على متن طائرة متجهة إلى بوسطن، وجمعت كذلك بيانات عينة مكونة من ٢٥٠ متأهلًا للعدو في ماراثون بوسطن. والآن لنفترض أن متوسط الوزن لكلا المجموعتين متساوٍ تقريبًا، فلنقل إنه ١٥٥ رطلًا. إن أيًّا منا ممن تزاحم في صف دخول طائرة مزدحمة، وجلس يتصارع على مسنَد مقعدها مع جاره في المقعد المجاور، يعرف يقينًا أن كثيرًا من ركاب الطائرات التجارية يزِنون أكثر من ١٥٥ رطلًا. لكنك كذلك تذكر أنه في تلك الرحلات المزدحمة البائسة أيضًا الكثير من الأطفال الرضع الصارخين والأطفال السيِّئي السلوك الذين يتمتعون بسعة رئوية تفوق أوزانهم بكثير. عندما يتم حساب متوسط الأوزان الموجودة على متن الطائرة، نجد وزن لاعبَي كرة القدم اللذين يزن الواحد منهما ٣٢٠ رطلًا، واللذان يجلسان على جانبَي مقعدك الأوسط، يعوضه ذلك الرضيع الضئيل الصارخ على الجهة الأخرى من الطائرة، أو ذلك الطفل ذو الستة سنوات والذي لا يفتأ يركل ظهر مقعدك.

طبقًا للأدوات الوصفية التي تعرفنا عليها حتى هذه اللحظة، فإن أوزان راكبي الطائرة ولاعبي الماراثون متطابقة تقريبًا. لكنها ليست كذلك. نعم وزن المجموعتين له نفس «الوسط» تقريبًا، لكن ركاب الطائرة أكثر تشتتًا حول نقطة الوسط تلك. فحتى ابني ذو الثمانية أعوام قد يلاحظ أن عدَّائي الماراثون يَبدون متساوي الوزن تقريبًا، بينما في الطائرة تجد أناسًا ضئيلي الجسد وآخرين ضخامًا بغرابة. إن أوزان ركاب الطائرة «أكثر تشتتًا»، وهي صفة مهمة في وصف المجموعتين. فالانحراف المعياري هو الإحصائية الوصفية التي تسمح لنا بالتعبير عن التشتت الحادث حول المتوسط من خلال رقم مفرد. سوف تجد معادلات حساب الانحراف المعياري والتباين (وهو مقياس عام آخر للتشتت وهو الذي قد اشتق منه الانحراف المعياري) واردة في ملحق في نهاية هذا الفصل. أما الآن فلنفكر في السبب وراء أهمية قياس التشتت.

لنفترض أنك ذهبت إلى عيادة الطبيبة، لأنك تشعر بالإعياء منذ أن تمت ترقيتك إلى منصب رئيس قسم جودة الطابعات في شمال أمريكا. تسحب الطبيبة عينة من دمائك وبعدها بأيام قليلة تجد رسالة صوتية على هاتفك من مساعدة الطبيبة تخبرك أن عدد الـ HCb2 في دمك (مادة كيميائية دموية مختلقة) يبلغ ١٣٤. حينها تهرع إلى الإنترنت فتكتشف أن متوسط عدد الـ HCb2 لشخص في عمرك هو ١٢٢ (وستجد الوسيط مثله تقريبًا). تبًّا، لو كنتَ مثلي لشرعت فورًا في كتابة وصيتك، ولكتبت – دامعًا – خطابات لوالديك وشريك حياتك وأبنائك وأصدقائك المقربين، بل لعلك أقدمت على تجربة القفز الحر من الطائرة، أو حاولت تأليف رواية بسرعة شديدة، ولأرسلت إلى رئيسك في العمل رسالة إلكترونية مشبِّهًا إياه بعضو معين من أعضاء الجسد البشري – بحروف كبيرة.

لكن أيًّا من هذه الأمور ليس ضروريًّا (وقد تكون لتلك الرسالة الإلكترونية عواقب وخيمة). فعندما ستهاتف عيادة الطبيبة كي تحجز لك مكانًا في قسم الحالات الميئوس منها، ستبلغك مساعدتها أن نتائج تحليلك تقع ضمن الحيز الطبيعي. لكن كيف يمكن هذا؟ سوف تشرع في الصياح بلا توقف في الهاتف «إن العدد يفوق المتوسط العادي باثنتي عشرة نقطة».

حينها ستجيبك تقنية المعمل قائلة باقتضاب: «إن الانحراف المعياري للـ HCb2 يبلغ ١٨ نقطة».

ما معنى هذا بحق الجحيم؟

هناك تنوع في عدد الـ HCb2 كما هو الحال مع معظم الظواهر البيولوجية (كالطول مثلًا). ففي حين يقدر متوسط تعداد جزيئات تلك المادة المزيفة بـ ١٢٢، فإن تعدادها لدى الكثير من الأشخاص الأصحاء قد يزيد أو ينقص عن هذا الرقم. ولا خطر هنا إلا إذا ارتفع الرقم أو انخفض بشكل متطرف. إذن كيف نستطيع أن نحدد «الشكل المتطرف»

في هذا السياق؟ كما عرفنا من قبل فإن الانحراف المعياري مقياس للتشتت، أي إنه يبين كيف تتجمع المشاهدات بشكل لصيق حول المتوسط. في الكثير من توزيعات البيانات المعتادة يقع جزء كبير من المشاهدات في داخل نطاق انحراف معياري واحد عن المتوسط (وهو ما يعني أنها تتدرج بين انحراف معياري تحت المتوسط وآخر فوقه). للشرح نورد هذا المثال البسيط؛ يقدَّر متوسط طول البالغين من الرجال الأمريكيين بخمس أقدام وعشر بوصات. ويقدر الانحراف المعياري عنه بنحو ثلاث بوصات، وهكذا فهناك شطر كبير من الرجال البالغين يتراوح طولهم ما بين خمس أقدام وسبع بوصات وبين ست أقدام وبوصة واحدة.

بتعبير آخر نقول إن أي رجل يقع طوله ضمن هذا النطاق لن يعتبر مفرط الطول أو قصيرًا بشكل غير معتاد. وهو ما يعيدنا مرة أخرى إلى نتائج تحليل الـ HCb2 غير المُطَمئنة. نعم عدد جزيئات تلك المادة الكيميائية في دمك يزيد عن المتوسط باثنتي عشرة نقطة، لكن هذا أقل من انحراف معياري واحد، والذي هو المكافئ لكون المرء يبلغ ست أقدام طولًا – أي إنه ليس بغير معتاد. بالطبع هناك بعض المشاهدات القليلة تقع على بُعد انحرافين معيارين من المتوسط، وأخرى أقل تقع على بُعد ثلاثة انحرافات معيارية أو أربعة. (في مثال الطول سيكون الأمريكي الذي يقع طوله على بُعد ثلاثة انحرافات معيارية من المتوسط يبلغ ست أقدام وسبع بوصات طولًا أو أكثر).

إن بعض التوزيعات أكثر تشتتًا من غيرها. ومن ثم فالانحراف المعياري لأوزان ركاب الطائرة الـ ٢٥٠ أعلى من نظيره لعدَّائي الماراثون الـ ٢٥٠. والتوزيع التكراري لأوزان ركاب الطائرة سيكون أكثر تخمة (بشكل حرفي) من التوزيع التكراري لعدَّائي الماراثون. ما إن نصل إلى المتوسط والانحراف المعياري حتى نجد أنفسنا أمام كشف معرفي كبير. فمثلًا افترض أنني قلت لك إن متوسط نتائج اختبار SAT * للرياضيات هو ٥٠٠ نقطة بانحراف معياري مقداره ١٠٠ نقطة. كما حدث مع مثال الطول، سوف يدخل جميع التلاميذ الذين خضعوا لهذا الاختبار ضمن نطاق انحراف معياري واحد عن المتوسط، أي ما بين ٤٠٠ و٦٠٠ نقطة. في رأيك كم من التلاميذ أحرز ٧٢٠ نقطة أو أكثر؟ عدد ليس بالكثير على الأرجح بما أن هناك أكثر من انحرافين معياريين يقع فوق المتوسط.

لكننا قادرون على أن نخرُج بنتيجة أفضل من «عدد ليس بالكثير»، فالآن فرصة سانحة لتقديم واحد من أهم التوزيعات في الإحصاء أهميةً وشيوعًا ونفعًا، ألا وهو التوزيع الاحتمالي الطبيعي. فالبيانات الموزعة بشكل طبيعي تكون متطابقة التموضع حول متوسطها في شكل يشبه الجرس سيبدو مألوفًا لك.

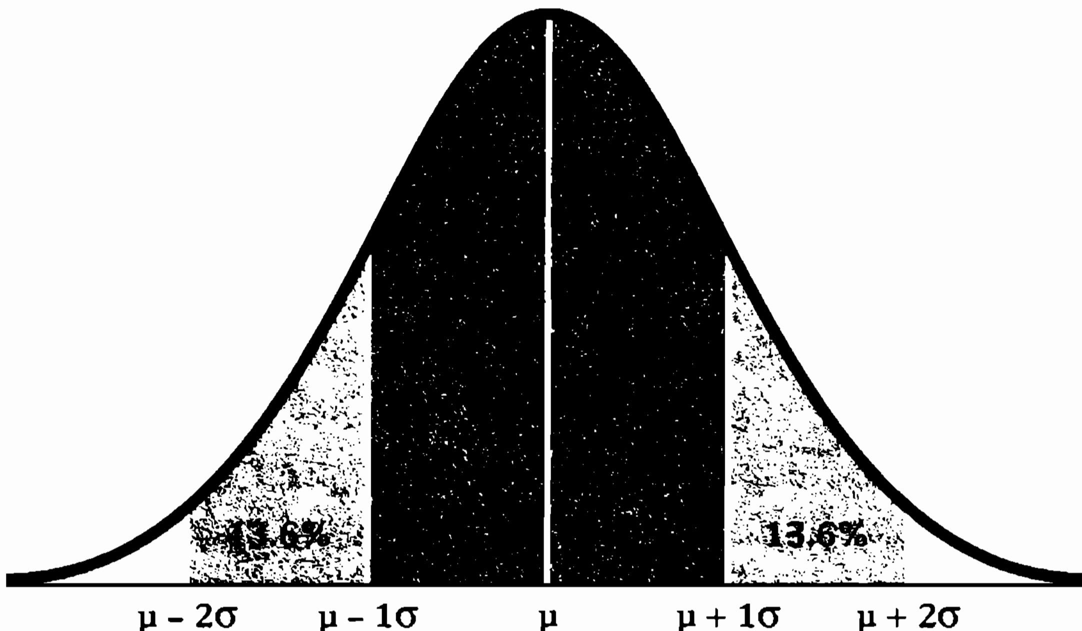
يصف التوزيع الاحتمالي الطبيعي العديدَ من الظواهر الشائعة. تخيل توزيعًا

* اختبار القبول في الجامعات الأمريكية وهو اختصار لعبارة Scholastic Aptitude Test. (المترجم).

تكراريًّا يصف الفشار المتفرقع على سطح الموقد. بعض حبات الفشار تتفرقع بسرعة، ربما تفرقع حبة أو اثنتان في كل ثانية، وفي غضون عشر أو خمس عشرة ثانية تصير حبات الفشار تتفجر بشكل محموم. ثم بمرور الثواني يأخذ عدد الحبات المتفرقعة يتناقص بالتدريج تقريبًا بنفس المعدل الذي بدأ به. إن أطوال الأمريكيين موزعة بشكل أكثر أو أقل من الطبيعي، بمعنى أنها موزعة بتناسقٍ بعض الشيء حول متوسطٍ مقدارُه خمس أقدام وعشر بوصات. وكل اختبار SAT مصمم تحديدًا لإنتاج توزيع احتمالي طبيعي لنتائج متوسطها ٥٠٠ نقطة وانحرافها المعياري ١٠٠ نقطة. وطبقًا لجريدة وول ستريت جورنال، فإن الأمريكيين يميلون إلى صفٍّ سياراتهم بتوزيع احتمالي طبيعي أمام المجمعات التجارية، فمعظم السيارات تُوقَف أمام مداخل المجمعات مباشرة – «قمة» المنحنى الطبيعي – مع «أذيال» من السيارات تمتد إلى يمين المدخل وإلى يساره.

إن جمال التوزيع الاحتمالي الطبيعي – فهو كمايكل جوردان في قوته وبراعته وأناقته – نابع من حقيقة أننا نعرف بالضبط أيُّ حصص المشاهدات الموزعة توزيعًا احتماليًّا طبيعيًّا تقع داخل نطاق انحراف معياري واحد عن المتوسط (٦٨,٢ بالمائة)، وأيُّها داخل نطاق انحرافين معياريين عن المتوسط (٩٥,٤)، وأيُّها داخل نطاق ثلاثة انحرافات معيارية (٩٩,٧)، وهلم جرًّا. قد تبدو لك المعلومات السابقة تافهة، ولكنها في الحقيقة الأساسُ الذي بُنيت عليه الكثير من الإحصائيات. في وقت لاحق سوف نعود لتلك النقطة بشكل أكثر عمقًا.

التوزيع الاحتمالي الطبيعي



ذلك الخط الأوسط الذي يمثّل دائمًا بالرمز اللاتيني μ هو المتوسط. أما الانحراف المعياري فيمثّل عادة بالرمز اللاتيني σ. وكل نطاق يمثّل انحرافًا معياريًّا واحدًا.

دائمًا ما تُستخدم الإحصائيات الوصفية للمقارنة بين عددين أو كميتين. إنني أفوق أخي طولًا ببوصة واحدة، درجة حراراة اليوم تزيد عن المتوسط التاريخي لهذا التاريخ، وهلم جرًّا. تبدو لنا هذه المقارنات منطقية لأن معظمنا يعرف مقياس الوحدات موضوع الدراسة. إن بوصة واحدة لا تُهم كثيرًا عندما يتعلق الموضوع بطول الشخص؛ لذا من هذا ستستنتج أنني وأخي بنفس الطول تقريبًا. لكن على العكس من ذلك تُعتبر تسع درجات انحرافًا كبيرًا على مستوى درجة الحرارة في ظل أي نوع من أنواع المناخ في أي وقت من أوقات السنة، فتسع درجات فوق المتوسط تجعل من أي يوم يومًا أكثر حرًّا من المعتاد. لكن لنفترض أنني قلت لك إن حبوب الجرانولا أ تفوق الجرانولا ب بواحد وثلاثين ميلليجرامًا من الصوديوم. ما لم تكن تعرف الكثير عن الصوديوم (وعن مقادير تقديم حبوب الجرانولا)، فلن تمنحك هذه العبارة الكثير من المعلومات. وماذا لو قلت لك إن قريبي آل قد نقصت أرباحه ٥٣٠٠٠ دولار عن مثيلتها العام الماضي؟ هل يعني هذا أن علينا أن نقلق على آل؟ أم إنه مدير محفظة وقائية والـ ٥٣٠٠٠ دولار تلك ما هي إلا خطأ تقريبي في أرباحه السنوية؟

لقد أغفلنا السياق في كلا المثالين السابقين. فأيسر الطرق لإضفاء مدلول على هاتين المقارنتين النسبيتين هي استخدام النسب المئوية. فإذا قلت لك إن قالب الجرانولا أ يحتوي على صوديوم أكثر بنسبة ٥٠ بالمائة من قالب الجرانولا ب، أو أن دخل العم آل قد انخفض بنسبة ٤٧ بالمائة العام الماضي، فسيكون لهذا الكلام معنًى. فعندما يعبَّر عن القياس بالنسب المئوية يمنحنا ذلك إحساسًا بالتدرج.

إنك على الأرجح قد تعلمت حساب النسب المئوية في الصف الرابع، ولسوف يغريك هذا بأن تتخطى الفقرات القليلة التالية. فليكن، ولكن لتقمْ بهذا التمرين البسيط لأجلي أولًا. لنفترض أن متجرًا كبيرًا يعرض فستانًا للبيع بسعر ١٠٠ دولار، ثم يقوم مساعد مدير المتجر بوضع خصم على كل البضاعة بمقدار ٢٥ بالمائة، لكن حدث بعدها أن طُرد مساعد المدير ذاك من العمل لأنه كان يتسكع في حانة مع بيل جيتس*، وقام المساعد الجديد بزيادة أسعار كل البضاعة بنسبة ٢٥ بالمائة. ما سعر الفستان الآن؟ إذا قلت (أو فكرت) إنه ١٠٠ دولار، فجدير بك ألا تتخطى أيًّا من فقراتي.

سيكون السعر النهائي للفستان ٩٣٫٧٥ دولارًا، وتلك ليست بالحيلة السحرية التي تجعل الجماهير تحييك وتتزلف إليك في حفلات الشراب. صحيح أن للنسب المئوية فوائد،

* جدير بالذكر هنا أن هذا الشخص كان واحدًا من العشرة ذوي الدخل السنوي المقدَّر بـ ٣٥٠٠٠ دولار، الذين كانوا يجلسون على مقاعد الحانة الدوارة عندما دخل عليهم بيل جيتس وببغائه الناطقة. يا لها من مفاجأة!

إلا أنها أحيانًا ما تكون مربكة بل وخادعة أيضًا. إن معادلة حساب فرق النسب المئوية (أو تغيرها) هي كما يلي: (العدد الجديد – العدد الأصلي) / العدد الأصلي. يعطينا البسطُ (الجزء الأعلى من الكسر) حجمَ التغير بقيم مطلقة، أما المقام (الجزء الأسفل من الكسر) فيضع ذلك التغير في سياق من خلال مقارنته بالنقطة التي بدأنا منها. في البداية قد تجد الأمر بسيطًا سهل الفهم تمامًا كما قام مساعد مدير المتجر بخصم ٢٥ بالمائة من سعر الفستان المقدر بـ ١٠٠ دولار. فخمسة وعشرون بالمائة من مائة دولار هي خمسة وعشرون دولارًا، وهذا هو الخصم الذي يخفض سعر الفستان إلى خمسة وسبعين دولارًا. يمكنك أن تُسقط الأرقام على المعادلة السابق ذكرها، ثم تقوم ببعض التلاعب البسيط كي تصل إلى نفس النقطة: (١٠٠ دولار – ٧٥ دولارًا) / ١٠٠ دولار= ٢٥ أو ٢٥ بالمائة.

حتى هذه اللحظة يُعرض الفستان بسعر خمسة وسبعين دولارًا إلى أن يأتي المساعد الجديد للمدير ويطلب رفع سعره بنسبة خمسة وعشرين بالمائة. وتلك هي النقطة التي يرتكب فيها معظم من يقرءون هذه الفقرة خطأً. حيث يتم حساب نسبة الخمسة والعشرين بالمائة من السعر الجديد المخفض للفستان، والذي هو خمسة وسبعون دولارًا. ستكون الزيادة ٠,٢٥ (٧٥ دولارًا)، أو ١٨,٧٥ دولارًا، وهكذا يصير السعر النهائي ٩٣,٧٥ دولارًا. (لا ١٠٠ دولار). المقصد هنا أنَّ تغير النسبة المئوية دائمًا يعطينا قيمة عددٍ ما نسبةً إلى شيء آخر. ومن ثم علينا أن نفهم ما هو ذلك الشيء الآخر.

قمت ذات مرة باستثمار قدر من المال في شركة أنشأها شريكي في السكن عندما كنت في الجامعة. ولأنها كانت مبادرة خاصة فلم تكن هناك اشتراطات للمعلومات التي تتاح لحاملي أسهم الشركة. مرت سنون دون أن أعرف أي معلومات عن مصير استثماري؛ إذ كان شريكي السابق في السكن كتومًا فيما يتعلق بهذا الموضوع. لكنني أخيرًا تلقيت خطابًا بالبريد يُعْلمني أن أرباح المؤسسة قد فاقت أرباح السنة الفائتة بنسبة ٤٦ بالمائة. لم تكن هناك أي معلومات عن تلك الأرباح بقيم مطلقة؛ بمعنى أنني ما زلت أجهل إذا ما كان استثماري رابحًا أم خاسرًا. إذا افترضنا أن المؤسسة قد ربحت في العام الماضي ٢٧ سنتًا، أي لا شيء فعليًّا، وربحت هذا العام ٣٩ سنتًا أي لا شيء أيضًا، لكن مع ذلك لقد نمت أرباح الشركة من ٢٧ سنتًا إلى ٣٩ سنتًا، فهذا يعني أنها ازدادت بنسبة ٤٦ بالمائة. من الواضح أن ذلك الخطاب الذي أُرسِلَ إلى حاملي الأسهم كان سيحبطهم لو ذُكر فيه أن الأرباح التراكمية للشركة على مدى عامين جاءت أقل قيمة من كوب قهوة من مقهى ستار باكس.

لأجل الإنصاف أقول إن شريكي بالسكن قد باع شركته في النهاية بمئات الملايين من الدولارات وأربحني مائة بالمائة من الأموال التي استثمرتها. (ولأنك لا تعرف كم من المال استثمرتُ، فلن تعرف كذلك كم ربحتُ – وهي الحقيقة التي تعزز فكرتي بطريقة لطيفة للغاية).

اسمح لي أن أضيف فارقًا آخر. لا ينبغي الخلط بين التغير في النسب المئوية والتغير في نقاط النسب المئوية. فعادةً ما يعبَّر عن المعدلات بالنسب المئوية. فمعدل ضريبة المبيعات يبلغ في إيلينوي ٦,٧٥ بالمائة. وإنني أدفع لوكيلي ١٥ بالمائة من عائدات كتبي. فهذه المعدلات تتم جبايتها على أساس بعض الكميات، كالدخل في حالة معدل ضريبة الدخل. من الواضح أن المعدلات تعلو وتنخفض بشكل لا يمكن التنبؤ به، ولهذا فإن التغيرات في المعدلات يمكن أن توصف من خلال طرائق شديدة التباين. وخير الأمثلة على ذلك التغير الذي طرأ حديثًا على ضريبة الدخل الشخصي في إيلينوي بارتفاعها من ٣ بالمائة إلى ٥ بالمائة. هنالك طريقتان للتعبير عن ذلك التغير في الضريبة، كلتاهما دقيقة من الناحية التقنية. فقد أشار الديمقراطيون (إشارة دقيقة) الذين خططوا هذا التغير الضريبي إلى أن معدل ضريبة الدخل في الولاية قد ارتفع نقطتين مئويتين (من ٣ بالمائة إلى ٥ بالمائة). أما الجمهوريون فقد أشاروا (إشارة دقيقة أيضًا) إلى أن ضريبة الدخل في الولاية قد ارتفعت بنسبة ٦٧ بالمائة. [ذلك اختبار عملي للمعادلة التي ذكرناها قبل بضع فقرات، وهي: (٥ – ٣)/ ٣ = ٢⁄٣، والتي تساوي ٦٧ بالمائة تقريبًا].

لقد ركز الديمقراطيون على التغير المطلق في المعدل الضريبي، أما الجمهوريون فقد ركزوا على تغير النسبة المئوية في العبء الضريبي. وكما ذكرنا فإن كلا الوصفين دقيق تقنيًّا، لكنني أزعم أن وصف الجمهوريين يبين بدقة أكثر تأثيرَ التغير الضريبي؛ حيث إن ما سأدفعه إلى الحكومة – المبلغ الذي يشغلني في مقابل طريقة حسابه – قد زاد بالفعل بنسبة ٦٧ بالمائة.

كثير من الظواهر تدحض توصيفات مثالية عن طريق إحصائية واحدة. افترض أن لاعب خط الوسط أرون رودجرز قد رمى الكرة لمسافات مجموعها ٣٦٥ ياردة، لكنه لم يصل لمنطقة إحراز الهدف ولا مرة واحدة. وفي نفس الوقت كان مجموع مسافات إلقاء الكرة من قِبل اللاعب بيتونماننج ١٢٧ ياردة فقط، لكنه نجح في إدخال الكرة لمنطقة إحراز الأهداف ثلاث مرات. لقد أحرز ماننج نقاطًا أكثر، أما رودجرز فقد هيَّأ إحراز الأهداف لفريقه من خلال تحريك فريقه في الملعب وإبعاد هجوم الفريق الآخر عن منطقتهم. مَن لعب بشكل أفضل؟ في الفصل الأول ناقشتُ معدلات التمريرات في البطولة الوطنية لكرة القدم، وهي المحاولة المنطقية التي قامت بها البطولة للتعامل مع ذلك التحدي الإحصائي. إن معدل التمريرات هو مثال لقائمة هي إحصائية وصفية مؤلفة من مجموعة من الإحصائيات الوصفية. ما إن تتم بلورة هذه المقاييس المختلفة في رقم واحد، حتى يتسنى استخدام هذه الإحصائية في عقد مقارنات على غرار وضع ترتيب لأفضلية لاعبي خط الوسط في يوم معين، أو حتى على مدى مسيرتهم الرياضية. لو كان في لعبة البيسبول قائمة مشابهة لتوافرت إجابة على السؤال المتعلق بأفضل اللاعبين على مر التاريخ. هل هذا حقيقي؟

إن ميزة أي قائمة هي أنها تبلور العديد من المعلومات المعقدة في رقم واحد. حينها نستطيع وضع تراتبية لأمور تتعصى على المقارنات البسيطة – أمور كلاعبي خط الوسط أو الكليات الجامعية أو المتسابقات على لقب ملكة الجمال. ففي مسابقة ملكة جمال أمريكا تكون نتيجة الفائزة باللقب هي مجموع لخمس منافسات منفصلة، وهي: المقابلة الشخصية والاستعراض بأثواب البحر، والاستعراض بثياب السهرة والموهبة والأسئلة التي تطرح عليهن على خشبة المسرح (أما ملكة اللطافة فتنتخب عبر اقتراع مستقل من قِبل المتسابقات أنفسهن).

لكن يا للخسارة، فعيب أي قائمة هي أنها تبلور الكثير من المعلومات المعقدة في رقم واحد. هناك طرائق من الكثرة أن تتعصى على العد لإنجاز تلك المهمة، وكلٌّ منها لها القدرة على الخروج بنتيجة مختلفة. يبين مالكولم جلادويل هذه الفكرة بطريقة رائعة في مقالة نشرها بصحيفة النيويورك تايمز، والتي انتقد فيها حاجتنا المُلحة لوضع تراتبية للأشياء[٢]. (وقد صب جامَّ غضبه على تراتبية الكليات الجامعية خصوصًا). يسوق جلادويل مثال ترتيب السيارة والسائق لثلاث سيارات رياضية هي: البورش كايمن، والشيفروليه كورفيت، واللوتس إيفورا. باستخدام معادلة تحتوي على واحد وعشرين متغيرًا، وضع ذلك المثال البورش في المرتبة الأولى. لكن جلادويل قد أشار إلى أن «الشكل الخارجي» للسيارة يقدر بـ ٤ بالمائة فقط من مجمل الترتيب في معادلة ذلك المثال، وهي النسبة التي تبدو منخفضة جدًّا بالنسبة إلى السيارات الرياضية. لو كان التصميم الخارجي قد منح ثقلًا أكبر في إجمالي الترتيب (٢٥ بالمائة) لتَصدَّرَته سيارة اللوتس.

لكن انتظر لحظة. لقد أشار جلادويل في مقالته إلى أن السعر المدوَّن على ملصق صانع السيارة له ثقل قليل نسبيًّا في معادلة ذلك المثال. إذا ما حظيت القيمة بثقل أكبر (مما يجعل الترتيب مرتكزًا بشكل متساوٍ على السعر والتصميم الخارجي وخصائص المركبة)، حينها ستكون الصدارة من نصيب الشيفروليه كورفيت.

تتسم القوائم كلها بالحساسية الشديدة تجاه الإحصائيات الوصفية التي تتضافر معًا لتكوينها، وكذلك تجاه الثقل الذي يحظى به كلُّ مكوِّن من هذه المكونات. ونتيجة لهذا نجد أمامنا مجموعة كبير من الأدوات تتدرج ما بين أدوات نافعة إلى أخرى زائفة كلية. ومن الأمثلة على ما سبق قائمة الأمم المتحدة للتنمية البشرية أو الـ HDI. وقد تم وضع هذه القائمة كمقياس للرفاهة الاقتصادية التي هي قائمة أوسع من أن تقتصر على الدخل. تستخدم تلك القائمة الدخل كواحد من مكوناتها، لكنها تضم أيضًا مقاييس المعدل العمري ومستوى التعليم. تحتل الولايات المتحدة المركز الحادي العشر على مستوى العالم بمقياس نصيب الفرد من الدخل (تأتي بعد أمم غنية بتروليًّا كقطر وبروناي والكويت)، لكنها تشغل المركز الرابع عالميًّا في مجال التنمية البشرية[٣]. صحيح أن قائمة الأمم المتحدة

للتنمية البشرية سوف تتغير قليلًا إذا ما تمت إعادة تهيئة الأجزاء المكونة للقائمة، لكن لن يحدث فيها أي تغير عقلاني يجعل زمبابوى تتقدم فتتخطى النرويج. تُقدِّم لنا هذه القائمة لمحة مفيدة تتسم بالدقة المنطقية عن المعايير المعيشية في مختلف بقاع العالم.

تمنحنا الإحصائيات الوصفية نظرة على الظواهر التي تجذب اهتمامنا. وبهذه الروح نستطيع أن نرجع إلى الأسئلة التي طرحناها في بداية هذا الفصل. من أفضل لاعبي البيسبول على مر التاريخ؟ لكن السؤال الأهم الذي يخدم غرض هذا الفصل هو، أيُّ الإحصائيات الوصفية أكثر فائدة في الإجابة عن هذا السؤال؟ طبقًا لكلام ستيف موير رئيس مؤسسة بيسبول إنفو سوليوشنز، فإن أكثر الإحصائيات الثلاث قيمةً (بخلاف العمر) في تقييم أي لاعب لا يلعب في مركز الرامي هي كما يلي:

١- النسبة القاعدية (OBP) وأحيانًا تسمى المتوسط القاعدي (OBA: وهي إحصائية تقيس مقدار الوقت الذي يصل فيه اللاعب إلى القاعدة بنجاح بما فيه المرات التي يمشي فيها (والتي لا تحسب ضمن متوسط ضرب الكرات).

٢- النسبة المئوية لقوة ضرب الكرات (SLG): تقيس قوة الضرب من خلال حساب إجمالي المرات التي وصلت فيها الكرة إلى المنطقة في مقابل مرات الضرب. يعبر عن المرة المفردة بالرقم ١، والمزدوجة بالرقم ٢، والثلاثية بالرقم ٣، وعن الضربة التي يكمل فيها اللاعب دورة كاملة ويعود إلى القاعدة دون إعاقة من الدفاع بالرقم ٤. وهكذا يكون الضارب الذي ضرب ضربة مفردة وأخرى ثلاثية في خمسة أدوار لعب له نسبة ضرب مئوية تقدر بـ (١+٣)/٥، أو ٨٠٠.

٣- معدل ضرب الكرات (AB): هي إحصائية تضع كل ما سبق في سياقه. فأي لاعب كان سيحظى بإحصائية مبهرة في مباراة أو اثنتين. أما «نجوم اللعبة» فيجمعون «الأرقام» المبهرة على مدى الآلاف من الأهداف المسجلة على لوحات الملاعب الإلكترونية.

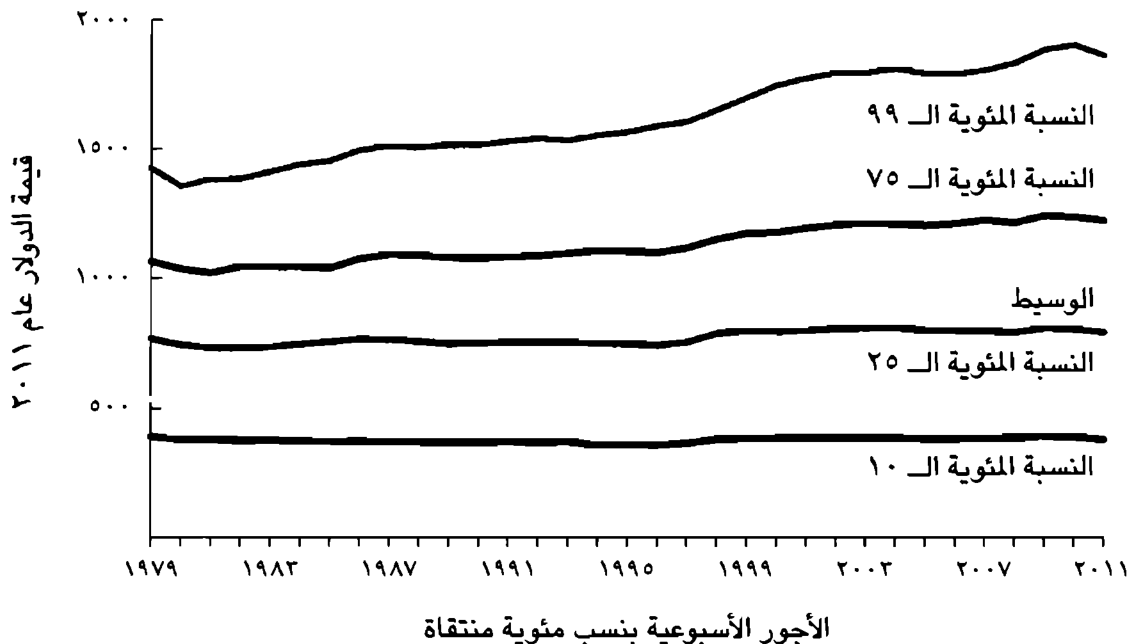
من وجهة نظر موير (والتي أضيفها بلا تردد) فإن أفضل لاعبي البيسبول على مر التاريخ هو بايب روث؛ وذلك لقدرته الفريدة على الضرب والرمي على السواء. ما زال بايب روث صاحب السجل الأعلى في مسيرته بالبطولة الكبرى بمعدل قوة ضرب الكرات تقدر بـ ٦٩٠ بالمائة[٤].

ماذا عن الصحة الاقتصادية للطبقة الوسطى الأمريكية؟ مرة أخرى ألجأ إلى الخبراء، فقد راسلت عن طريق البريد الإلكتروني كلًّا من جيف جروجر (أحد زملائي في جامعة شيكاجو) وآلان كروجر (نفس عالم الاقتصاد الذي درس ظاهرة الإرهاب والذي يعمل حاليًّا رئيس مجلس المستشارين الاقتصاديين للرئيس أوباما). وكلاهما أجابني نفس الإجابة بطريقة مختلفة. لتقييم الصحة الاقتصادية للطبقة الوسطى الأمريكية، فعلينا أن نفحص التغير الحادث في الأجر الوسيط (المعدل لأجل التضخم) على مدى بضعة عقود.

كما اقترح عليَّ الزميلان تفحُّص التغيرات التي تطرأ على الأجور بنسبة ٢٥ و٧٥ بالمائة (والتي يمكن أن تترجم منطقيًّا إلى الحدين الأدنى والأعلى للطبقة الوسطى).

لكن لا يزال هناك فارق آخر. فعندما نقوم بتقييم الصحة الاقتصادية سنتفحص إما الدخل وإما الأجور فليسا بالشيء نفسه. فالأجر هو ما نتقاضاه في مقابل مقدار ثابت من العمل، كالأجر عن الساعة أو الأجر الأسبوعي. أما الدخل فهو مجموع كل ما نتقاضاه من مصادر مختلفة. فإذا ما عمل أشخاص بوظيفة ثانية أو عملوا لساعات أطول فسيزيد دخلهم لكن أجرهم لن يتغير. (بهذا الشكل قد يرتفع الدخل حتى لو انخفض الأجر إذا ما افترضنا أن العامل ينفق وقتًا كافيًا في وظيفته). لكن برغم ذلك فإذا ما اضطر العامل إلى مزيد من العمل طلبًا لمزيد من المال، فسوف يكون من الصعب تقييم أثر ذلك على رفاهته. فالأجر مقياس أقل إبهامًا فيما يتعلق بالكيفية التي يعوض الأمريكيون بها عن الأعمال التي يضطلعون بها، فكلما زاد الأجر زاد ما يتقاضاه العاملون عن كل ساعة يقضونها في وظائفهم.

بعد كل ما سبق دعني أقدم لك رسمًا بيانيًّا للأجور الأمريكية خلال العقود الثلاثة الأخيرة، كما أنني أضفت النسبة المئوية الـ ٩٠ لتبيان التغييرات التي طرأت على أجور من ينتمون للطبقة الوسطى مقارنة على مدى ذلك الإطار الزمني بأولئك العاملين الذين يتصدرون ذلك التوزيع.



الأجور الأسبوعية بنسب مئوية منتقاة

المصدر: «Changes in the Distribution of Workers' Hourly Wages between 1979 and 2009»، Congressional Budget Office، February 16، 2011. بيانات هذا الرسم البياني موجودة في موقع: http://www.cbo.gov/sites/default/files/cbofiles/ftbdocs/120xx/doc12051/02-16-wagedispersion.pdf.

يمكننا أن نستقي من هذه البيانات مجموعة متنوعة من الخلاصات. إنها لا تقدم إجابة مفردة «صحيحة» تتعلق بالرفاهة الاقتصادية للطبقة الوسطى، وإنما تخبرنا بأن العامل العادي أو العامل الأمريكي الذي يتقاضى الأجر الوسيط كان كمن «يجري في مكانه» لنحو ثلاثين عامًا. أما العاملون الذين يتقاضون النسبة الـ ٩٠ بالمائة فحالهم أفضل بكثير. إن الإحصائيات الوصفية يمكن أن تساعدنا على تأطير المسألة. أما الإجراء الذي سنتخذه حيال الأمر إذا ما فعلنا فهذا سؤال أيديولوجي وسياسي.

ملحق الفصل الثاني

جداول بيانات أعطال الطابعات

	صفر	واحد	اثنان	ثلاثة	أربعة	خمسة	ستة	سبعة	ثمانية	تسعة	عشرة أو أكثر
تكرار الأعطال لدى الشركة المنافسة	12	14	36	13	8	6	5	3	0	2	1
	صفر	واحد	اثنان	ثلاثة	أربعة	خمسة	ستة	سبعة	ثمانية	تسعة	عشرة أو أكثر
تكرار أعطار شركتك	25	31	9	4	3	0	0	1	1	0	26

معادلة التباين والانحراف المعياري

إن التباين والانحراف المعياري هما أكثر آليتين إحصائيتين لوصف وقياس تشتت التوزيع شيوعًا. والتباين الذي يرمز إليه عادة بالرمز σ^2 يتم حسابه عن طريق تحديد كم يبعد موقع المشاهدات الحادثة على بُعد توزيع واحد عن المتوسط. لكن الحيلة تكمن هنا في تربيع كل فارق بين كل مشاهدة والمتوسط، ثم تتم قسمة مجموع كل تلك القيم المربعة على عدد المشاهدات.

أو على وجه التحديد:

لكل مجموعة من مشاهدات $x_1, x_2, x_3 \ \ x_n$ بمتوسط μ

التباين يساوي $\sigma^2 = \left[(x_1-\mu)^2+(x_2-\mu)^2+(x_3-\mu)^2+\cdots(x_n-\mu)^2\right]/n$

لأن الفارق بين كل مصطلح ومتوسطه مربع، فإن معادلة حساب التباين تضع ثقلًا معتبرًا على المشاهدات التي تقع بعيدًا عن المتوسط، أو التي تسمى بالقيم المتطرفة، كما في الجدول التالي الذي يبين أطوال التلاميذ.

المجموعة 1	الطول (μ = 70) بوصة	المسافة من المتوسط = القيمة المطلقة لـ $(x_n - \mu)$*	$(x_n - \mu)^2$	المجموعة 2	الطول (μ = 70) بوصة	المسافة من المتوسط = القيمة المطلقة لـ $(x_n - \mu)$*	$(x_n - \mu)^2$
نيك	74	4	16	سحر	65	5	25
إيلانا	66	4	16	ماجي	68	2	4
دينا	68	2	4	فيصل	69	1	1
ربيكا	69	1	1	تيد	70	0	0
بن	73	3	9	جيف	71	1	1
تشارو	70	0	0	نارسيسو	75	5	25
		المجموع = 14	المجموع = 46			المجموع = 14	المجموع = 56
			التباين = =6/46 7,7				التباين = = 6/56 93
			الانحراف المعياري = $\sqrt{7.7} = 2.8$				الانحراف المعياري = $\sqrt{9.3} = 3$

يبلغ متوسط الطول في كلتا المجموعتين ٧٠ بوصة. كما أن أطوال التلاميذ في كلتا المجموعتين تفرق عن المتوسط بنفس العدد الإجمالي للبوصات ألا وهو ١٤. وطبقًا لمقياس التشتت هذا، فإن المجموعتين متطابقتان. لكن رغم ذلك نجد التباين في المجموعة الثانية أعلى بسبب الثقل الموضوع في معادلة التباين على القيم التي تقع بعيدة عن المتوسط – سحر ونارسيسو في هذه الحالة.

نادرًا ما يستخدم التباين في حد ذاته كإحصائية وصفية. وإنما تتجلى فائدته في أفضل صورها عند استخدامه كخطوة من خطوات حساب الانحراف المعياري لتوزيعٍ ما، والذي هو أداة أكثر بداهة تستخدم كإحصائية وصفية.

* القيمة المطلقة هي المسافة بين رقمين، بصرف النظر عن اتجاه تلك المسافة، بحيث تظل تلك القيمة موجبة دائمًا. في هذه الحالة تمثل تلك القيمة عدد البوصات بين طول الشخص والمتوسط.

الانحراف المعياري لمجموعة من المشاهدات هو الجذر المربع للتباين:

لكل مجموعة من مشاهدات $\chi_1, \chi_2, \chi_3 \ \ \chi_n$ بمتوسط μ

الانحراف المعياري = σ = الجذر المربع لكل تلك الكمية =

$$\sqrt{\left[(x_1-\mu)^2+(x_2-\mu)^2+(x_3-\mu)^2+\cdots(x_n-\mu)^2\right]/n}$$

الفصل الثالث

وصف خادع

كعبارة «إنه شخص ذو شخصية رائعة!» وغيرها من العبارات الحقيقية مع كونها مضللة في مجملها

إن المتفكرين في المواعدة والعلاقات يجدون عبارة «إنه شخص ذو شخصية رائعة» عبارة تدق في أذهانهم أجراس إنذار. وليس هذا لأن العبارة كاذبة بالضرورة، وإنما لأجل ذلك الذي لا تفصح عنه، كحقيقة أن ذاك الشخص صاحب سوابق جنائية، أو أن إجراءات طلاقه لم «تنتهِ بعد». إننا لا نشك في كون ذلك الشخص ذا شخصية رائعة، وإنما نخشى أن هذا القول الحق قد قيل لإخفاء معلومات أخرى بطريقة شديدة التضليل (على افتراض أن معظمنا لا يحبذ مواعدة محكومين سابقين ما زالوا متزوجين). ليست تلك العبارة كذبة في حد ذاتها، بمعنى أنها لا تعتبر شهادة زور، لكنها ما زالت من غير الدقة أن صارت لا تخبر بالحقيقة.

وكذلك هو الإحصاء. فرغم أن مجال الإحصاء هو مجال متجذر في الرياضيات، والرياضيات علم دقيق، فإن استخدام الإحصاء لوصف الظواهرالمعقدة ليس بالأمر الدقيق. وهو ما يترك مجالًا كبيرًا للتعتيم على الحقيقة. لمارك تواين عبارة شهيرة ذكر فيها أن للأكاذيب ثلاثة أنواع؛ وهي: الأكاذيب، والأكاذيب اللعينة، والإحصائيات*. وكما اتضح في الفصل السابق فإن معظم الظواهر التي تشغل تفكيرنا يمكن أن توصف بطرائق عدة. ولما كانت هناك طرائق عدة لوصف نفس الشيء (كعبارتي «إنه ذو شخصية رائعة»، أو «لقد أدين في قضية تزوير أوراق مالية» مثلًا)، فإن الإحصائيات الوصفية التي نختار استخدامها (أو عدم استخدامها) ستحدث أثرًا كبيرًا في الانطباع الذي نتركه على الآخرين. فقد يَستخدم شخص ذو دوافع خبيثة حقائق وأرقامًا صحيحة تمامًا كي يدعم نتائج خبيثة أو غير مشروعة.

---

* نَسَب تواين هذه العبارة لرئيس الوزراء البريطاني بنجامين دزرائيلي، لكن ليس هناك مِن توثيق يؤكد أن دزرائيلي قد كتب هذه العبارة أو قالها.

علينا أن نبدأ في التفريق الجوهري بين «الدقة» و«التحديد». فهاتان الكلمتان لا تحل إحداهما محل الأخرى. فالدقة تدل على الإحكام الذي نعبِّر به عن شيءٍ ما. فإذا ما وصفتَ طول المسافة التي تقطعها خلال انتقالك فإن مقولة «٤١,٦ ميلًا» أكثر دقة من «نحو أربعين ميلًا»، والتي هي بدورها أكثر دقة من مقولة «مشوار طويل نجس». وإذا ما سألتني: كم تبعد أقرب محطة وقود عن هنا؟ فقلت لك إنها على بُعد ١,٢٦٥ ميلًا شرقًا، فتلك إجابة دقيقة. لكن المشكلة هنا هي أن نفس المقولة ستكون غير محددة بشكل تام إذا ما كانت المحطة في الاتجاه المعاكس. لكنني إذا ما قلت لك: «واصل القيادة لعشر دقائق أو نحوها حتى تصل إلى منصة بيع نقانق، ستجد محطة الوقود على بُعد بضع مئات الياردات بعدها إلى يمينك. إذا ما وجدت نفسك قد تخطيت مطعم الهوترز فقد ابتعدتَ كثيرًا»، فإجابتي تلك أقل دقة من عبارة «١,٢٦٥ تجاه الشرق»، لكنها أفضل منها بكثير؛ لأنني أوجِّهك إلى جهة محطة الوقود. فالتحديد هو مقياس لمدى تطابق العدد مع الحقيقة – وهو ما يوضح خطر الخلط بين الدقة والتحديد. فإذا ما كانت إجابةٌ ما دقيقةً فسيكون مزيد من التحديد أفضل. لكن مهما زاد قدر التحديد فإنه لا يغني أبدًا عن الدقة.

بل الواقع أن الدقة يمكن أن تخفي عدم تحديد من خلال إعطائنا شعورًا زائفًا باليقين سواء عن عمد أو غير عمد. في عام ١٩٥٠ بلغ جوزيف ماكارثي السيناتور الأمريكي عن ولاية ويسكونسن، والمغالي في استخدام سياسة الاتهام الأحمر*، ذروة حملته الخرقاء في توجيه الاتهامات، عندما لم يكتفِ بادعاء أن الشيوعيين قد اخترقوا وزارة الخارجية الأمريكية، وإنما أيضًا أنه يمتلك قائمة بأسماء أولئك المخترقين. فأثناء خطابه في ويلينج بفرجينيا الغربية لوَّح ماكارثي بقطعة من الورق وأعلنها قائلًا: «ها هي في يدي قائمة مؤلفة من ٢٠٥ أسماء يعلم وزير الخارجية أنهم أعضاء بارزون في الحزب الشيوعي، لكنهم رغم هذا لا يزالون يعملون ويرسمون ملامح سياسة وزارة الخارجية»[١]. فيما بعد اتضح أنه لم تكن بالورقة أية أسماء، لكن دقة الاتهام أكسبته مصداقية رغم أنها كانت محض كذبة فجة.

أما أنا فقد أدركت أهمية التفرقة بين الدقة والتحديد من سياق أقل خبثًا. فذات سنة أهدتني زوجتي في عيد الميلاد جهاز تحديد نطاق خاصًا بلعبة الجولف يقوم بحساب المسافات بين كرة الجولف والحفرة. يعمل هذا الجهاز عن طريق نوع من أنواع أشعة الليزر، فأقف أنا قرب كرتي على الممر المهد (أو الوعر)، وأصوب الجهاز تجاه العلم المزروع قرب الحفرة، عندها يحسب لي الجهاز بالضبط المسافة التي يتعين عليَّ ضرب الكرة لتقطعها. يعتبر هذا الجهاز تطورًا كبيرًا عن علامات تحديد الياردات التي لا تبين

---

* الاتهام الأحمر هو مصطلح اصطُلِحَ في عشرينيات القرن الماضي الماضي، ويُقصد به دحض أفكار الخصم من خلال اتهامه بالشيوعية (المترجم).

إلا المسافة التي تفصل تلك البقاع عن مركز الملعب (مما يجعلها دقيقة لكن غير محددة). بالاستعانة بهدية عيد الميلاد صرت أعرف أنني أبعد عن الحفرة بمسافة ١٤٧,٢ ياردة. كنت أتوقع أنْ تُحسِّن دقة التكنولوجيا الأنيقة بهذا الجهاز من لعبي للجولف، لكنها في الواقع قد زادتها سوءًا.

وجدت أمامي مشكلتين؛ أولاهما أنني ظللت أستخدم ذلك الجهاز الغبي لثلاثة أشهر كاملة قبل أن أدرك أنه مهيأ للقياس بالأمتار لا الياردات، وهكذا كان كل حساب يبدو دقيقًا كـ (١٤٧,٢) حسابًا خطأً. وثانيتهما أنني كنت أحيانًا أصوب الشعاع دون قصد تجاه الأشجار التي تقع خلف المسطح الأخضر، لا تجاه العلم المشير إلى الحفرة، وهكذا كانت ضربتي «المثالية» تقطع المسافة المفترضة منها بالضبط عابرة المسطح الأخضر ثم مستقرة في الغابة. كان الدرس الذي تعلمته من هذه التجربة، وهو الدرس الذي ينطبق على جميع التحليلات الإحصائية، هو أنه حتى أكثر القياسات أو الحسابات دقة لا بد أن تتماشى مع المنطق العام.

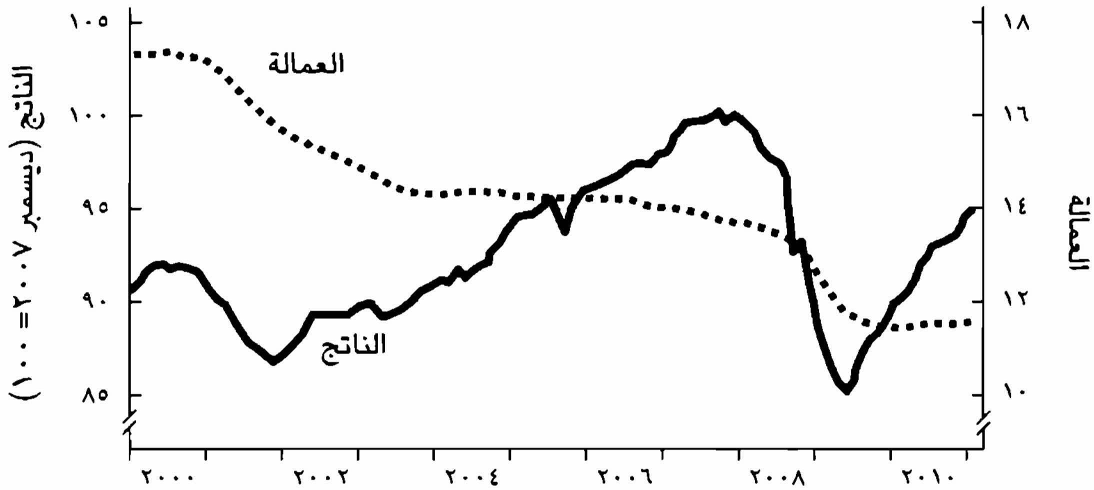
لنَسُقْ مثالًا ذا تبعات أكثر فداحة؛ نقول إن العديد من تقييمات المخاطر في وول ستريت، والتي سبقت الأزمة المالية عام ٢٠٠٨، كانت دقيقة جدًّا. فمفهوم «القيمة المهددة بالخطر» قد سمح لبعض المؤسسات أن تقدر بدقةٍ حجمَ رءوس أموال كل مؤسسة، والتي يحتمل أن تضيع طبقًا لسيناريوهات مختلفة. تمثلت المشكلة في أن تلك النماذج المنمقة المعقدة قد فعلت فِعل من هيَّأوا جهاز تحديد النطاق للقياس بالأمتار لا بالياردات. كانت الحسابات الرياضية معقدة غامضة، لكن الإجابات التي كانت تخرج بها كانت دقيقة بشكل مطمئن. أما عن الفرضيات التي طُرحت بشأن ما قد يحدث للأسواق العالمية المندمجة في تلك النماذج، فقد كانت خطأً تمامًا، مما جعل تلك النتائج غير محددة في إجمالها بطرق لم تسبب اضطراب وول ستريت فقط، وإنما جعلت الاقتصاد العالمي بأسره يضطرب.

حتى أكثر الإحصائيات الوصفية دقة وتحديدًا يمكن أن تعيبها مشكلة جوهرية، ألا وهي: الافتقار إلى الوضوح حول ما هو مطلوب تعريفه أو وصفه أو شرحه. والجدالات الإحصائية تشبه كثيرًا الزيجات التعيسة؛ فالمتخاصمون فيها دائمًا ما يقاطعون بعضهم كلام بعض. دعنا نتأمل هذا السؤال الاقتصادي المهم: إلى أيِّ مدًى تُعتبر الصناعة الأمريكية متمتعة بعافيتها؟ عادة ما نسمع عن أن قطاع الصناعة الأمريكي يخسر الكثير من الوظائف لصالح الصين والهند وغيرهما من البلدان ذات الأجور المنخفضة. ونسمع كذلك أن التصنيع العالي التكنولوجيا ما زال مزدهرًا في الولايات المتحدة الأمريكية، وأن أمريكا تظل واحدة من أكبر البلاد المصدرة للبضائع المصنعة. أيُّ هاتين المقولتين أصدق؟ قد تبدو لك هذه واحدة من الحالات التي قد يعيننا فيها تحليل سليم ذو بيانات مضبوطة على حسم

الجدل بين الروايتين. هل يربح مجال التصنيع في الولايات المتحدة وينافس عالميًا، أم إنه ينكمش أمام المنافسة الأجنبية؟

الواقع أن العبارتين صحيحتان. فقد جمعت المجلة الإخبارية البريطانية الإيكونوميست بين هاتين العبارتين المتناقضتين، اللتين تصفان حال مجال التصنيع في أمريكا، من خلال الرسم البياني التالي:

«تعافي منطقة تصنيع الصلب» ١٠ مارس ٢٠١١



هذا التناقض الظاهري منبعه الكيفية التي نعرف بها «الصحة» في مجال الصناعة الأمريكية. فمن منظور النتائج – أي القيمة الإجمالية للبضائع المنتجة والمَبيعة – نما القطاع الصناعي في الولايات المتحدة بشكل مضطرد خلال العشرية الأولى من القرن الحالي، ثم تلقَّى ضربة كبيرة خلال فترة الركود العظيم، ثم بعدها نهض من عثرته واستمر حتى الآن ينمو مرة أخرى. يتوافق هذا مع بيانات كتاب الحقائق العالمية الذي أصدرته وكالة المخابرات المركزية الأمريكية، والذي أظهر أن الولايات المتحدة هي ثالث أكبر مصدِّر في العالم بعد الصين وألمانيا. لا تزال الولايات المتحدة واحدًا من مراكز القوة التصنيعية.

لكن الرسم البياني الذي أوردته الإيكونوميست به خط ثانٍ، وهو خط العمالة في مجال التصنيع. خلال العقد الأخير استمر عدد الوظائف في مجال الصناعة بالولايات المتحدة في الانكماش؛ إذ فقدَ نحو ستة ملايين شخص وظائفهم. هاتان القصتان معًا – ارتفاع ناتج التصنيع وانخفاض معدلات العمالة – تخبرانِنا بالحكاية كاملة. لقد نما قطاع التصنيع في الولايات المتحدة باضطراد وصار أكثر إنتاجية، بمعنى أن المصانع صارت تنتج أكثر بعدد أقل من العمال. يعد هذا أمرًا جيدًا فيما يتعلق بالتنافسية العالمية؛ حيث إنها تجعل المنتجات الأمريكية أكثر قدرة على منافسة بضائع البلدان ذات الأجور المنخفضة. (واحدة من الطرق للتنافس مع مؤسسة تدفع للعامل دولارين عن ساعة العمل

الواحدة هي خلق عملية تصنيع من الفعالية أن تجعل عاملًا واحدًا يتقاضى أربعين دولارًا عن الساعة الواحدة قادرًا على تكرارها لمرات تفوق مرات العامل ذي الدولارين عشرين مرة). لكن الوظائف في مجال التصنيع قد تقلصت جدًّا، وهو أمر فظيع بالنسبة إلى أولئك العمال المسرَّحين الذين يعتمدون على أجورهم من ذلك المجال.

بما أن هذا الكتاب يتعلق بالإحصاء لا الصناعة، فلنرجع إلى النقطة الأساسية، والتي هي أن «صحة» قطاع الصناعة الأمريكي – والتي هي أمر يسهل تقدير حجمه من الناحية الظاهرية – تعتمد على الطريقة التي نختار أن نعرف بها مفهوم الصحة: الناتج أم العمالة؟ في تلك الحالة (وحالات كثيرة غيرها) نصل إلى أكثر القصص اكتمالًا عندما نجمع بين الرقمين، تمامًا كما تصرفت الإيكونوميست بحكمة فأوردت الرسم البياني السابق.

حتى إذا ما اتفقنا على مقياس واحد للنجاح، كاختبارات التلاميذ الدراسية، تظل هناك فرصة كبيرة للمناورات الإحصائية. انظر إذا ما كنت قادرًا على أن توفِّق بين العبارتين الافتراضيتين التاليتين، واللتان يمكن أن تكونا حقيقيتين:

السياسي أ (الطامح إلى الكرسي): «إن مدارسنا في انحدار! فقد انخفضت معدلات المدارس في نتائج الاختبارات الدراسية عن العام الماضي بنسبة ستين بالمائة».

السياسي ب (الذي يحتل الكرسي حاليًّا): «إن مدارسنا في تقدم! فثمانون بالمائة من تلاميذنا قد أحرزوا نتائج أعلى من العام الماضي».

إليك هذه الملحوظة: لا تحتوي كل المدارس بالضرورة على نفس عدد التلاميذ. إذا ما تأملت العبارتين السابقتين المتناقضتين ظاهريًّا مرة أخرى لوجدت أحد السياسيين قد استخدم المدارس كوحدة تحليل «ستين بالمائة من المدارس ...»، بينما استخدم الآخر التلاميذ كوحدة تحليل «ثمانون بالمائة من التلاميذ ...». ووحدة التحليل هي كيان يقارن أو يوصف عن طريق الإحصاء – كأداء المدارس وأداء التلاميذ. فمن الممكن جدًّا أن يتحسن أداء التلاميذ بينما ينحدر أداء معظم المدارس – إذا ما كان التلاميذ الذين يظهر عليهم التحسن يدرسون في مدارس كبيرة الحجم جدًّا. ولجعل هذا المثال أكثر منطقية دعنا نطبقه على الولايات الأمريكية:

السياسي أ (شعبوي): «إن اقتصادنا قد آل إلى الحضيض! فقد هبط دخل ثلاثين ولاية العام الماضي».

السياسي ب (نخبوي نوعًا ما): «إن اقتصادنا قد حقق مكاسب معتبرة، فقد ارتفعت دخول سبعين بالمائة من الأمريكيين العام الماضي».

قد نستنتج من هاتين العبارتين أن أكبر الولايات تتمتع بأكثر الاقتصاديات صحة، كما هو الحال في نيويورك وكاليفورنيا وتكساس وإيلينوي وغيرها. أما الثلاثون ولاية

ذات متوسطات الدخل الآخذة في الانخفاض فهي أصغر حجمًا على الأرجح، كما هو الحال في فرمونت ونورث داكوتا ورود آيلاند وغيرها. وهذا التفاوت في حجم الولايات يعزز احتمالية كون غالبية الولايات الأمريكية تعاني اقتصاديًّا في حين يشهد الأمريكيون انتعاشًا اقتصاديًّا. الدرس الأهم هنا هو أن ننتبه لوحدة التحليل. أن نسأل سؤال «من» نَصِف؟ أو «ماذا» نَصِف؟ وكيف يفرق عمن أو ما يصفه غيرنا؟

مع كون الأمثلة السابقة هي أمثلة افتراضية، إليك هذا السؤال الإحصائي الجوهري غير الافتراضي: هل زادت العولمة من تفاوت الدخل حول العالم أم قلَّصته؟ أحد التفسيرات ينص على أن العولمة لم تفعل سوى أن فاقمت التفاوت في الدخل؛ فالدول التي كانت أكثر ثراءً عام ١٩٨٠ (كما حددها نصيب الفرد من الدخل) اتجهت لأن تنمو بين عامي ١٩٨٠ و٢٠٠٠ بسرعة تفوق سرعة البلدان الأفقر[٢]. ما حدث هو أن البلدان الغنية قد ازدادت ثراءً، مما دعَّم فكرة أن التجارة والاستعانة بمصادر خارجية والاستثمارات الأجنبية، وغيرها من مكونات «العولمة»، ما هي إلا أدوات يستخدمها العالم المتقدم لبسط هيمنته الاقتصادية. فلتسقط العولمة! فلتسقط العولمة!

لكن تمهل للحظة. فنفس تلك البيانات يمكن (ويجدر بها) أن تفسر بشكلٍ مختلفٍ كليةً إذا ما غيرنا وحدة التحليل. إننا لا نبالي بالدول الفقيرة وإنما نبالي بالفقراء من الناس. وإن شطرًا كبيرًا من فقراء العالم يعيشون في الصين والهند. وهم دولتان ضخمتان (يتجاوز تعداد الواحدة منهما المليار نسمة) وكلتاهما كانتا فقيرتين نسبيًّا عام ١٩٨٠. لقد شهدت الهند والصين نموًّا متسارعًا على مدى العقود الماضية، وكان جزءٌ كبيرٌ من هذا النمو مَرَدُّه إلى التكامل الاقتصادي المتزايد مع بقية العالم. إنهما كما وصفتهما الإيكونوميست «دولتان سريعتا التعولم». ولأن هدفنا هو رفع المعاناة الإنسانية، فليس من المنطقي أن نساوي بين الصين (ذات الـ ١٫٣ مليار نسمة) وموريشيوس (ذات الـ ١٫٣ مليون نسمة) ثم نشرع في تفحُّص آثار العولمة على الفقراء.

يجب أن تكون وحدة التحليل ممثَّلة في البشر لا الدول. وما حدث فعلًا بين عامي ١٩٨٠ و٢٠٠٠ يشابه كثيرًا مثال اختبارات المدارس المختلق الذي أوردتُه سابقًا. وغالبية فقراء العالم تَصادَف أنهم يعيشون في دولتين عملاقتين نَمَتَا بسرعة مع اندماجهما في الاقتصاد العالمي؛ لذا فإن التحليل المناسب سوف يأتي بنتائج مغايرة تمامًا حول الفوائد التي يجنيها فقراء العالم من العولمة. أو كما أشارت الإيكونوميست قائلة: «إذا ما نظرتَ إلى الشعوب لا الدول لوجدتَ التفاوت في الدخل على مستوى العالم آخِذًا في الانخفاض بشكل متسارع».

مؤخرًا نشب صراع حملات إعلانية بين شركتَي الاتصالات إيه تي آند تي وفيريزون، استغلت فيه كل شركة منهما نفس ذلك الإبهام الذي يكتنف ما تصفان. كلتا الشركتين

تقدم خدمة الهاتف المحمول. واحد من الأمور التي يهتم بها مستخدمو الهاتف المحمول هو مقدار جودة الخدمة في الأماكن التي عادةً ما يستخدمون الهاتف فيها لإجراء المكالمات أو تلقِّيها. ومن ثم قد يبدو أن نقطة المقارنة المنطقية بين الشركتين هو حجم شبكتيهما وجَوْدَتيهما. ولما كان المستهلكون لا يهتمون إلا بالحصول على خدمة جيدة في الكثير من الأماكن، فقد وضعت الشركتان، كل على حدة، مقاييس مختلفة لتقدير ذلك الطلب الهلامي على «الخدمة الجيدة في الكثير من الأماكن». أطلقت شركة فيريزون حملة إعلانية هجومية تروِّج لما حدث من تحوُّل جغرافي في شبكتها، لعلك تذكر الخرائط التي عرضوها للولايات المتحدة، والتي أظهرت النسبة الكبيرة من الدولة التي تغطيها شبكة فيريزون مقارنةً بالقدر الضئيل جغرافيًّا الذي تغطيه شبكة إيه تي آند تي. لقد اختار فيريزون الرقعة الجغرافية المغطاة كوحدة تحليل، وذلك بسبب تفوُّقها في ذلك.

لكن شركة إيه تي آند تي ردَّت الهجوم بإطلاق حملة غيَّرت فيها وحدة التحليل. فقد وضعت لوحات إعلانية ضخمة قالت فيها «إنَّ إيه تي آند تي يَستخدم خدمتها ٩٧ بالمائة من الأمريكيين». لاحظ هنا استخدام كلمة «الأمريكيين» بدلًا من «أمريكا». لقد اهتمت إيه تي آند تي بحقيقة أن غالبية الأمريكيين لا يعيشون في المنطقة الريفية بمونتانا أو في صحراء أريزونا. فبما أن السكان لا يتوزَّعون بشكل متساوٍ على مستوى الرقعة الجغرافية للولايات المتحدة الأمريكية بأسرها، فإن السبيل إلى خدمة الهاتف المحمول الجيدة (وهو ما تهدف إليه الحملة بشكل ضمني) هو امتلاك شبكة فعالة في الأماكن التي يعيش الناس ويعملون، وليس بالضرورة في الأماكن التي يذهبون إليها للتخييم. وبصفتِي شخصًا يقضي وقتًا كبيرًا في منطقة نيو هامبشاير الريفية، فإنني لا بد أن أتعاطف مع فيريزون في هذا الأمر.

صديقانا القديمان، المتوسط والوسيط، يمكن أن يُستخدما كذلك لغايات غير حميدة. لعلك تذكر من الفصل السابق أن الوسيط والمتوسط، كلاهما مقياس «لوسط» التوزيع أو «نزعته إلى المركزية». والمتوسط هو معدل بسيط: إجمالي المشاهدات مقسومًا على عددها. (متوسط ٣ و٤ و٥ و٦ و١٠٢ هو ٢٤). أما الوسيط فهو نقطة المنتصف في التوزيع؛ حيث تقع نصف المشاهدات فوق الوسيط، ويقع النصف الآخر تحته. (وسيط ٣ و٤ و٥ و٦ و١٠٢ هو ٥). هنا سيلاحظ القارئ الحاذق، الفارق المعتبر بين ٢٤ و٥. فإذا ما رغبتُ، لسببٍ ما، في وصف مجموعة الأرقام تلك بطريقة تُضَخِّمها، فسوف أركز على المتوسط. أما إذا أردتُ أن أجعلها تبدو صغيرة، فسوف أركز على الوسيط.

والآن دعنا ننظر إلى تطبيق هذه الفكرة في العالم الواقعي. لنتأمل تخفيضات الضرائب التي قام بها جورج دبليو. بوش، وهو المشروع الذي روَّجت له إدارة بوش كمشروع يصب في صالح غالبية الأسر الأمريكية. وفي إطار تلك الحملة الترويجية للخطة،

ذكرت الإدارة أن ٩٢ مليون أمريكي سوف تتقلص الضرائب التي يدفعونها بما يزيد عن ١٠٠٠ دولار (١٠٨٣ دولار على وجه التحديد). لكن هل يتسم ذلك التلخيص لخفض الضرائب بالدقة؟ تقول صحيفة النيويورك تايمز: «إن البيانات لا تكذب لكن بعضها صامت».

هل ستتقلص ضرائب ٩٢ مليون أمريكي؟ نعم.

هل سيبلغ مقدار تقلص الضرائب لدى هؤلاء الناس ١٠٠٠ دولار؟ لا، سوف يبلغ وسيط انخفاض الضرائب أقل من ١٠٠ دولار.

مجموعة صغيرة نسبيًّا مؤلفة من الأفراد فاحشي الثراء هي فقط المؤهلة لأن تحظى بتخفيضات ضريبية كبيرة، وهذه الأرقام الكبيرة تسبب انحراف المتوسط، جاعلةً معدل الانخفاض الضريبي يبدو أكبر مما قد يتمتع به معظم الأمريكيين. فالوسيط ليس حساسًا تجاه القيم المتطرفة، وهو في هذه الحالة يعد وصفًا أكثر دقة، على الأرجح، للكيفية التي يؤثر بها الانخفاض الضريبي على الأسرة الأمريكية العادية.

لكن الوسيط كذلك يمكن أن يكون غير صريح بشكل كامل؛ وذلك لأنه ليس حساسًا تجاه القيم المتطرفة. افترض أنك تعاني من مرض قد يودي بحياتك، لكنك تلقيت أخبارًا جيدة مفادها أن عقَّارًا جديدًا قد ظهر من الممكن أن يكون فعالًا. لكن الخبر السيء هو أن هذا العقَّار باهظ الثمن وأنه ذو آثار جانبية سيئة كثيرة. حينها تسأل: «لكن هل سيشفيني؟». يخبرك الطبيب أن العقَّار الجديد سيزيد من وسيط فترة حياتك عن بقية المرضى الذين يعانون نفس مرضك بمقدار أسبوعين. هذه أخبار لا تكاد تكون سعيدة، فيبدو أن هذا العقَّار لا يستحق تكاليفه ولا ما يصاحبه من متاعب. حينها سترفض شركة التأمين أن تدفع تكلفة العلاج استنادًا إلى حجة قوية تعتمد على رقم وسيط فترة بقائك على قيد الحياة.

لكن الوسيط قد يمثل إحصائية مضللة جدًّا في هذه الحالة. لنفترض أن كثيرًا من المرضى لم يستجيبوا للعلاج الجديد بينما تَسبَّب هذا العلاج في شفاء عدد كبير من المرضى بما يعادل ٣٠ أو ٤٠ بالمائة. لن يتضح هذا النجاح من الوسيط (رغم أن متوسط فترة بقاء أولئك الذين تناولوا العقَّار على قيد الحياة سيبدو رائعًا). في هذه الحالة ستكون القيم المتطرفة – ممثَّلة في أولئك الذين تعاطوا العقَّار وعاشوا لفترة أطول – مرتبطة جدًّا بالقرار الذي ستتخذه. وهذه ليست مجرد حالة افتراضية. إذ يُذكر أن عالم الأحياء التطورية ستيفن جاي جولد قد شُخِّصَت حالته بأحد أنواع السرطانات، وحُدِّدَ وسيط فترة بقائه على قيد الحياة بثمانية أشهر، لكنه توفي بعد عشرين عامًا بسبب نوع آخر لا علاقة له بالنوع الأول[٢]. كان جولد قد كتب مقالًا شهيرًا عنوانه «الوسيط ليس برسالة». وفيه أكد أن معرفته العلمية بالإحصاء أنقذته من تلك النتيجة الخطأ التي قضت بوفاته في ظرف

ثمانية أشهر. فتعريف الوسيط يخبرنا أن نصف المرضى سيعيشون لثمانية أشهر على الأقل – وربما أكثر من ذلك بكثير جدًّا. في هذه الحالة يكون توزيع نسب الوفيات «منحرفًا نحو اليمين»، وهو الأمر الذي سيعني لك الكثير إذا ما كنت تعاني هذا المرض[4].

في هذا المثال نجد الصفة المميزة للوسيط – أنه لا يزن المشاهدات طبقًا لمدى بُعدها عن نقطة المنتصف وإنما فقط طبقًا لموقعها في الأعلى أو الأسفل – اتضح أنها نقطة ضعفه. أما المتوسط على النقيض من ذلك فيؤثر عليه تشتُّته. فيما يتعلق بالدقة، نجد التقابل بين المتوسط والوسيط يتمحور حول ما إذا كانت القيم المتطرفة تشوِّه ما يوصف، أم إنها تعتبر جزءًا مهمًّا من الرسالة. (مرة أخرى يتغلب المنطق على الرياضيات). بالطبع لن نقول إن المرء مضطر لأن يختار بين المتوسط والوسيط. فأي تحليل إحصائي شامل سيقدِّم الاثنين على الأرجح. وإذا ما قدَّم واحدًا منهما فقط فغالبًا ما سيكون هذا طلبًا للإيجاز، أو ربما لأن أحدهم يريد أن «يقنع» الآخرين مستخدمًا الإحصاء.

إن المُعمَّرين منا قد يذكرون الحوار التالي (كما أذكره)، والذي دار بين تشيفي تشايس وتيد نايت في فيلم كادي شاك. يلتقي الرجلان في غرفة خلع الملابس بعد خروجهما للتو من ملعب الجولف:

تيد نايت: كم ضربت؟

تشيفي تشايس: لا أسجل نقاطي.

تيد نايت: إذن كيف تقارن نفسك ببقية اللاعبين؟

تشيفي تشايس: طبقًا لطول القامة.

لن أحاول أن أفسر لك الفكاهة في الحوار السابق، وإنما سأقول لك إن الكثير من المناورات الإحصائية تنشأ من المقارنات على طريقة «تفاحات وبرتقالات». لنفترض أنك تحاول أن تقارن بين سعر غرفة في فندق بلندن وأخرى في باريس. حينها ستجعل ابنتك التي هي في السادسة من عمرها تفتح جهاز الكمبيوتر وتبحث في الأمر على شبكة الإنترنت لأنها أسرع منك وأكثر قدرة. ستقول لك ابنتك إن غرف الفنادق في باريس أغلى ثمنًا منها في لندن؛ فالغرفة في باريس تكلف ١٨٠ في الليلة، بينما تكلف في لندن ١٥٠ في الليلة الواحدة.

حينها ستشرح لطفلتك الفرق بين الجنيه الإسترليني واليورو، ثم ستعيدها إلى الكمبيوتر مرة أخرى كي تبحث لك عن فرق سعر الصرف بين العملتين حتى يتسنى لك عقد مقارنة منطقية. (هذا المثال يعتمد نوعًا ما على واقعة حقيقية؛ حيث إنني بعد أن دفعت ١٠٠ روبية هندية في مقابل ملء إبريق من الشاي في الهند، سألتني ابنتي عن سبب غلو أسعار كل شيء في هذه الدولة). فبالطبع لا تعني الأرقام المدونة على عملات الدول

المختلفة شيئًا إلا بعد أن نحوِّلها إلى وحدات قابلة للمقارنة. ما فرق سعر الصرف بين اليورو والجنيه الإسترليني، أو في حالة الهند بين الروبية والدولار؟

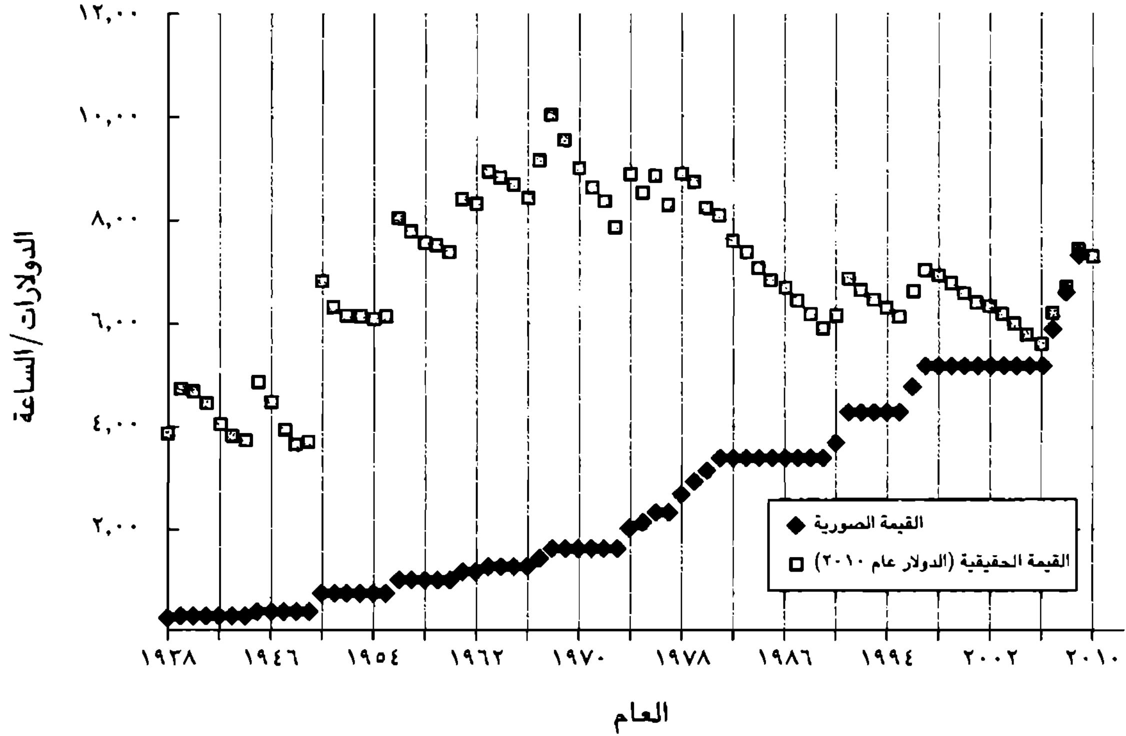
يبدو هذا درسًا واضحًا بشكل مزعج، إلا أنه في نفس الوقت درس يتجاهله الناس دائمًا، وخاصة السياسيين واستديوهات هوليوود. إن هؤلاء الناس - ولا شك - يدركون الفرق بين اليورو والجنيه الاسترليني، لكنهم مع ذلك يتغافلون عن مثال آخر من أمثلة التفاحات والبرتقالات، وإن كان أكثر مخاتلة، ألا وهو التضخم. فالدولار اليوم ليس كالدولار قبل ستين عامًا، فهو يشتري أشياء أقل بكثير. فبسبب التضخم صار الشيء الذي كان ثمنه دولارًا واحدًا عام ١٩٥٠ يكلف ٩,٣٧ دولارًا عام ٢٠١١. ونتيجة لهذا فإن أي مقارنة مالية بين عامي ١٩٥٠ و٢٠١١ دون اعتبار التغيرات التي طرأت على قيمة الدولار، ستكون أقل دقة من مقارنة أسعارٍ باليورو والجنيه الإسترليني - بما أن اليورو والجنيه الإسترليني تقترب قيمتهما إحداهما من الأخرى أكثر من اقتراب قيمة الدولار عام ١٩٥٠ منها عام ٢٠١١.

تلك ظاهرة مهمة حتى إن الاقتصاديين يحرصون على الإشارة إلى ما إذا كانت الأرقام قد عُدلت طبقًا للتضخم أم لا. لكن الأرقام الصورية لا تُعدَّل طبقًا للتضخم؛ فالمقارنة بين التكلفة الصورية للبرنامج الحكومي عام ١٩٧٠ والتكلفة الصورية لنفس البرنامج عام ٢٠١١ ليست إلا مقارنة بين الشيكات التي أصدرتها الحكومة في هذين العامين، دون أي اعتبار لحقيقة أن الدولار عام ١٩٧٠ كانت له قوة شرائية تفوق نظيره عام ٢٠١١. فإذا ما كنا قد صرفنا عشرة ملايين دولار عام ١٩٧٠ على برنامج يهدف لتزويد قدامى المحاربين بمعونات إسكانية، وأنفقنا أربعين مليون دولار على نفس البرنامج عام ٢٠١١، فمعنى هذا أن التزام الحكومة الفدرالية بهذا البرنامج قد انخفض مستواه. صحيح أن معدل الإنفاق قد ارتفع من الناحية الصورية، لكن هذا لم يعكس التغير في قيمة الدولارات التي أُنفقت. فدولار واحد عام ١٩٧٠ يساوي ٥,٨٣ دولارات عام ٢٠١١، وهكذا تحتاج الحكومة لأن تنفق ٥٨,٣ مليون دولار على إسكان قدامى المحاربين عام ٢٠١١ كي تعادل الملايين العشرة التي أنفقتها عام ١٩٧٠

أما الأرقام الحقيقية فهي على النقيض من الأرقام الصورية من حيث كونها معدلة طبقًا للتضخم. وفي هذا الصدد تعتبر أكثر المنهجيات نجاحًا هي تحويل جميع الأرقام إلى وحدة واحدة، كسعر الدولار عام ٢٠١١ مثلًا، في سبيل عقد مقارنات من نوع «تفاحات وتفاحات». هناك مواقع إلكترونية كثيرة على شبكة الإنترنت، من بينها موقع إحصائيات مكتب العمل الأمريكي، تحوي حاسبات تضخم بسيطة تقوم بمقارنة قيم الدولار في نقاط زمنية مختلفة*. إذا ما أردت أن ترى مثالًا حقيقيًّا (نعم أنا ألعب بالكلمات) للكيفية التي

* متوفر على موقع: http://www.bls.gov/data/inflation_calculator.htm.

يتغير بها مظهر الإحصائيات، إذا ما هيئت طبقًا للتضخم، فانظر لذلك الرسم البياني التالي الذي يمثل الحد الفدرالي الأدنى للأجور في الولايات المتحدة، والذي يبين القيمة الصورية والقيمة الحقيقية للقوة الشرائية كلتيهما للدولار عام ٢٠١٠.



المصدر: http://oregonstate.edu/instruct/anth484/minwage.html

إن الحد الأدنى للأجور – ذلك الرقم المدوَّن على لوحة الإعلانات في ركن قصيٍّ بمكتبك – قد حدده الكونجرس. ذلك الأجر الذي يبلغ حاليًّا ٧,٢٥ دولارات هو رقم صوري. لا يتعين على رئيسك في العمل أن يتأكد من أن الـ ٧,٢٥ دولارات تلك تشتري نفس ما كانت تشتريه قبل سنتين مضتا، وإنما كل ما يتعين عليه، في الواقع، هو أن يتأكد من أنك تتقاضى ذلك المبلغ عن كل ساعة تعملها.

لكن بمرور الزمن يقلص التضخم من القوة الشرائية للحد الأدنى للأجور (وغيره من الأجور الصورية، وهو السبب الذي يجعل النقابات لا تفتأ تناقش «التأقلم مع تكاليف المعيشة»). فإذا ما ارتفعت الأسعار بوتيرة أسرع من وتيرة رفع الكونجرس للحد الأدنى للأجور، حينها ينخفض المدفوع مقابل ساعة العمل. وعلى من يحددون هذا الحد الأدنى من الأجور أن يُعْنُوا بالقيمة الحقيقية لذلك الأجر بما أن المقصد الأساسي من هذا القانون هو ضمان حصول من يتقاضون هذا الحد الأدنى على مستوًى أدنى من القوة الشرائية عن كل ساعة عمل، لا أن يُعْنُوا بمنحهم شيكات مدوَّنًا عليها أرقام كبيرة لم تعُد تشتري ما كنت تشتريه سابقًا. (لو كان هذا هو المقصد لدفعنا لأصحاب الحد الأدنى للأجور بالروبية الهندية).

تعتبر استديوهات هوليوود من أكثر المتغافلين، بشكل فاضح، عن تلك الانحرافات التي يسببها التضخم عند مقارنة الأرقام في فترات زمنية مختلفة – وعن عمد أيضًا. ما الأفلام الخمسة التي كانت الأكثر جَنْيًا للأرباح (على المستوى المحلي) على مدى تاريخ السينما عام ٢٠١١؟[٥]

آفاتار (٢٠٠٩).

تيتانيك (١٩٩٧).

الفارس الأسود (٢٠٠٨).

حرب النجوم الحلقة الرابعة (١٩٧٧).

شريك ٢ (٢٠٠٤).

عندما تنظر في تلك القائمة قد تساورك بعض الريبة. نعم كانت تلك أفلامًا ناجحة – لكن شريك ٢؟ هل حقق ذلك الفيلم نجاحًا تجاريًّا فاق فيلم ذهب مع الريح؟ أو العراب؟ أو الفك المفترس؟* كلا وكلا ثم كلا. يروق لهوليوود أن تجعل كل فيلم ضخم يبدو أكثر نجاحًا من سابقه. واحدة من طرق عمل ذلك هي كتابة عناوين الأخبار المتعلقة بشباك تذاكر الأفلام محوِّلين الأرباح إلى الروبية الهندية على غرار ما يلي: «فيلم هاري بوتر يكسر الرقم القياسي في شباك التذاكر بعد أن جنى في عطلة نهاية الأسبوع ١,٣ تريليون». لكن حتى أقل مرتادي السينما ذكاءً سيرتابون من هذه الأرقام؛ لأنها محولة لعملة ذات قوة شرائية محدودة نوعًا ما. لكن استديوهات هوليوود (والصحفيين الذين يكتبون عنها) تكتفي بدلًا من هذا باستخدام قيم صورية، وهو الأمر الذي يجعل أحدث الأفلام تبدو كما لو أنها قد نجحت نجاحًا ساحقًا لأن أسعار التذاكر أعلى مما كانت عليه قبل عشرة أو عشرين أو خمسين عامًا. (عندما بدأ عرض فيلم ذهب مع الريح في دور العرض عام ١٩٣٩ كان سعر التذكرة وقتها يدور في فلك الـ ٠,٥٠ دولار أمريكي). إن أكثر الطرق دقة في مقارنة النجاح التجاري على مر الزمن هي إعادة تهيئة الأسعار طبقًا للتضخم. فجنْي أرباح تقدر بـ ١٠٠ مليون دولار عام ١٩٣٩ هو أمر أكثر إثارة للإعجاب من جني ٥٠٠ مليون دولار عام ٢٠١١. إذن ما الأفلام الخمسة الأعلى من حيث النجاح التجاري في الولايات المتحدة على مر الزمن؟[٦]

ذهب مع الريح (١٩٣٩)

حرب النجوم الحلقة الرابعة (١٩٧٧)

صوت الموسيقى (١٩٦٥)

إي. تي. (١٩٨٢)

* الاسم التجاري الذي عُرف به فيلم Jaws في الوطن العربي (المترجم).

الوصايا العشر (١٩٥٦)

وهكذا صار فيلم آفاتار يحتل المركز الرابع عشر فعليًّا، أما شريك ٢ فيهبط إلى المركز الحادي والثلاثين.

لكن حتى عند المقارنة بطريقة التفاحات في مقابل التفاحات، لا تزال الفرص متاحة لحدوث تضليل. كما ذكرنا في الفصل السابق، فإن واحدًا من أهم الأدوار التي يقوم بها الإحصاء يتمثل في وصف التغيرات التي تطرأ على الكميات بمرور الزمن. هل تتصاعد الضرائب؟ كم نبيع من شطائر البرجر بالجبنة مقارنة بالعام الماضي؟ إلى أي حد قللنا من الزرنيخ الموجود في مياه الشرب التي نتجرعها؟ عادةً ما نستخدم النسب المئوية للتعبير عن تلك التغيرات لأنها تمنحنا حسًّا بالقياس والسياق

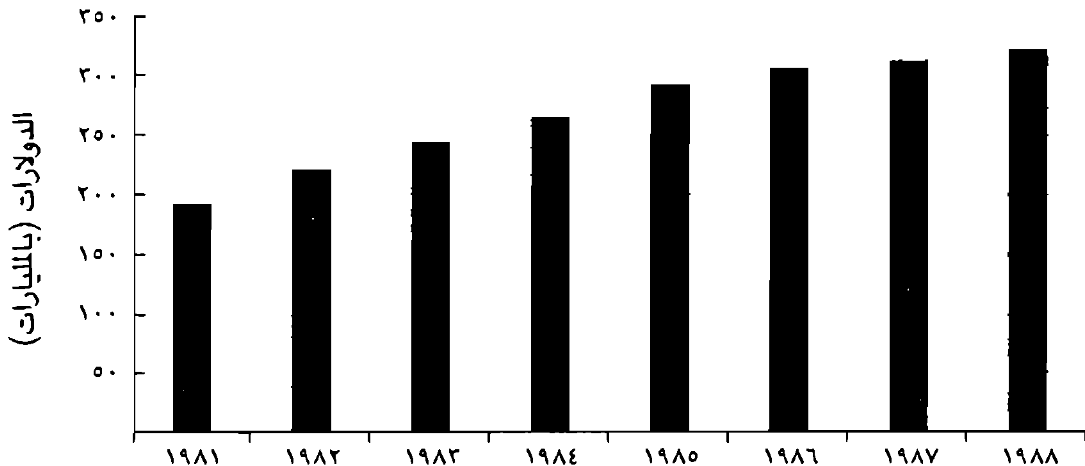
فإننا قادرون على فهم معنى خفض نسبة الزرنيخ بنسبة ٢٢ بالمائة، في حين لا يفهم منا إلا القليل إذا ما كان تقليل الزرنيخ بمقدار ميكروجرام واحد (والذي هو التقليل المطلق) يعد إجراءً مهمًّا أم لا. فالنسب المئوية لا تكذب – لكنها تبالغ أحيانًا. فواحدة من الطرق التي تجعل النمو السكاني يبدو كانفجار سكاني تتمثل في وصف تغيُّرٍ ما، نسبةً إلى نقطة بداية منخفضة جدًّا. إنني أعيش في مقاطعة كوك بولاية إيلينوي، وذات يوم هالني خبر مفاده أنه قد تَقرَّر رفع حصة ضرائبي المستقطعة بنسبة ٥٢٧ بالمائة لصالح دعم مصحة السل بضاحية مقاطعة كوك، لكنني مع ذلك قد ألغيت ذلك الحشد المناهض للضرائب الذي كنت أُزمِع إنشاءه (والذي كان لا يزال في طور التخطيط) بعدما علمت أن ذلك التغيير سيكلفني ما يقل عن ثمن شطيرة لحم ديك رومي. إن مقاطعة مصحة مرض السل تتعامل مع نحو مائة حالة مرضية سنويًّا، وهي ليست بمؤسسة ضخمة ورسومها ليست باهظة. أشارت صحيفة شيكاجو صن تايمز إلى أن الضرائب التي يدفعها المواطن العادي من مُلَّاك المنازل سوف ترتفع من ١,١٥ دولار إلى ٦ دولارات[٧]. أحيانًا ما يتحفظ الباحثون على رقم معين ممثل للنمو، موضحين أنه «ناشئ عن قاعدة منخفضة»؛ بمعنى أن أي تغير عن تلك القاعدة سيبدو كبيرًا عند المقارنة.

من الواضح أن هذا الجانب الملتوي حقيقي، فالنسبة المئوية الصغيرة من مجموع ضخم قد تكون رقمًا كبيرًا. لنفترض أن وزير الدفاع قد قال إن ميزانية الإنفاق على الدفاع سوف ترتفع بنسبة ٤ بالمائة هذا العام. هي أخبار عظيمة إذن! لا ليست كذلك لأن ميزانية وزارة الدفاع تقارب الـ ٧٠٠ مليار دولار. ونسبة ٤ بالمائة من ٧٠٠ مليار تساوي ٢٨ مليار دولار أمريكي، وهو المبلغ الذي قد يبتاع عددًا كبيرًا من شطائر لحم الديك الرومي. بل إن تلك الزيادة التي قد تبدو تافهة لكونها تبلغ ٤ بالمائة فقط، هي في الواقع تتجاوز كامل ميزانية وكالة ناسا وتتساوى مع ميزانيتَي وزارتَي العمل والخزانة مجتمِعتين.

وفي نفس الإطار تجد رئيسك في العمل الطيب القلب يوضح لك أنه من قبيل العدالة سوف يتقاضى جميع الموظفين نفس العلاوة هذا العام، وستكون ١٠ بالمائة من الراتب. يا لها من لفتة كريمة عدا أن رئيسك ذاك إن كان يتقاضى مليون دولار سنويًّا وتتقاضى أنت ٥٠ ألف عن نفس المدة، فستبلغ علاوته ١٠٠٠٠٠ بينما لن تزيد أنت إلا ٥٠٠٠ دولار في السنة. وعبارة «الكل سيزيد راتبه بنسبة ١٠ بالمائة هذا العام» أفضل وقعًا على الآذان بكثير من عبارة «ستكون علاوتي أكبر من علاوتك بعشرين مِثْلًا»، رغم أن العبارتين كلتيهما حقيقيتان في هذه الحالة.

لا بد أن يكون لأي مقارنة للتغير الكمي على مر الزمن نقطةُ بداية ونقطةُ نهاية. لكن المرء قادر على أن يتلاعب بتلك النقاط بطرائق تؤثر في الرسالة التي يرغب في إيصالها. واحد من أساتذتي في الجامعة كان يحب أن يتكلم عن «صور الحزب الجمهوري المسلطة» و«صور الحزب الديمقراطي المسلطة». كان يشير إلى بيانات الإنفاق على وزارة الدفاع، وما عناه هنا أنه كان يستطيع إعادة ترتيب نفس البيانات بطريقة تَسُرُّ الديمقراطيين، ثم يرتبها بطريقة أخرى فتَسُرُّ الجمهوريين. فهو عندما يخاطب الجمهوريين يقدم لهم الصورة التالية، والتي بها بيانات عن زيادة الإنفاق على الدفاع إبان حكم رونالد ريجان. لا يخفى على أحد أن ريجان قد ساعد على استعادة التزامنا بالدفاع والأمن، وهو الأمر الذي ساعدنا على الانتصار في الحرب الباردة. لا أحد يستطيع أن يطالع تلك الأرقام ثم لا يجد في نفسه تقديرًا لذلك العزم الفولاذي لرونالد ريجان على مواجهة السوفييت.

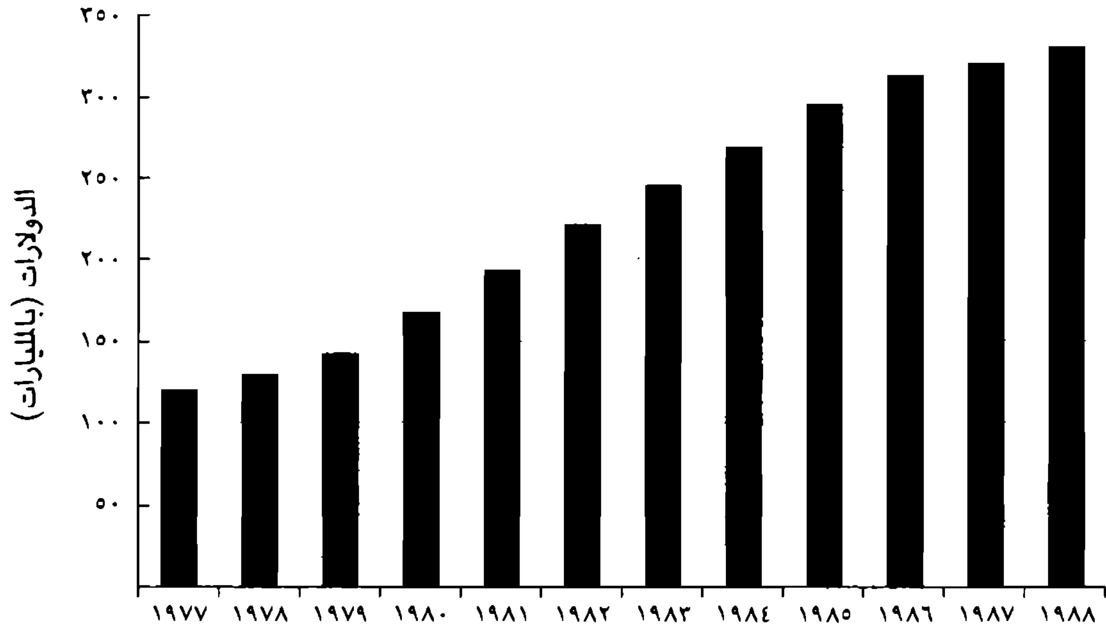
الإنفاق على الدفاع بالمليارات ١٩٨١ – ١٩٨٨



لكن أستاذي السابق كان حين يخاطب الديمقراطيين يستخدم ذات البيانات تقريبًا (الصورية) ولكن يقوم بإطالة الإطار الزمني. فقد أشار لتلك المجموعة إلى أن جيمي كارتر له الفضل في استهلال هذه القوة الدفاعية التراكمية. وكما تظهر تلك الصورة

«الديمقراطية»، فالارتفاع الذي حدث في الإنفاق الدفاعي ما بين عامي ١٩٧٧ و١٩٨٠ يُظهر نفس نمط الارتفاع الذي حدث خلال رئاسة ريجان. الشكر للرب لأنه جعل جيمي كارتر – خريج أكاديمية أنابوليس البحرية وضابط البحرية السابق – يبدأ في جعل أمريكا قوية مرة أخرى.

الإنفاق على الدفاع بالمليارات ١٩٧٧ – ١٩٨٨



المصدر: http://www.usgovernmentspending.com/spend.php?span=usgs302&year=1988&view=1&expand=&units=b&fy=fy12&local=s&state=US&pie=#usgs302.

صحيح أن الغاية الأساسية من الإحصائيات هي تقديم صورة تحمل معنًى عن تلك الأمور التي نهتم بها، لكننا في العديد من الحالات نرمي إلى أن نتخذ إجراءات قائمة على تلك الإحصائيات. فالفِرَق المشاركة في البطولة الوطنية لكرة القدم الأمريكية تريد مقياسًا بسيطًا لأداء لاعبي خط الوسط حتى يتسنى لهم إيجاد الموهوبين من لاعبي الجامعات وضمُّهم إلى صفوفهم. والمؤسسات تريد قياس أداء موظفيها حتى ترقِّي ذوي القيمة منهم وتفصل من هم دونهم. في عالم الأعمال يتردد قول مأثور نصه: «إنك لا تستطيع أن تدير ما لا تستطيع قياسه». هذا صحيح، لكن عليك أن تتوثق جيدًا من أن ما تقيسه هو بالفعل ما تحاول إدارته.

لنتناول موضوع جودة المدارس. إنه موضوع محوري نحتاج إلى قياسه إذا رغبنا في مكافأة المدارس «الجيدة» واستنساخها وكذلك معاقبة المدارس «السيئة» وإصلاحها. (وفي داخل كل مدرسة من تلك المدارس نجد نفس ذلك التحدي بالنسبة إلى قياس كفاءة

المدرسين لنفس السبب الأساسي). ويعتبر أكثر المقاييس شيوعًا بالنسبة إلى المدارس والمدرسين على السواء هو نتائج الاختبارات. فإذا ما حقق التلاميذ نتائج مبهرة في اختبارات موحدة محكمة الصياغة، فإن المدرسة والمدرسين يُبلون بلاءً حسنًا على الأرجح. والعكس صحيح؛ إذ تشير النتائج السيئة إلى الحاجة إلى فصل كثير من الناس بشكل عاجل. تستطيع هذه الإحصائيات أن تقودنا عبر طريق طويل نحو إصلاح نظام التعليم العام لدينا، أليس كذلك؟

لا ليس الأمر كذلك، فأي تقييم للمدرسين والمدارس يعتمد فقط على نتائج الاختبارات سوف يخرج بصورة غير دقيقة لحد الخطر. فالتلاميذ الذين يرتادون مدارس مختلفة تتنوع قدراتهم ويأتون من خلفيات متغايرة. فنحن نعلم، على سبيل المثال، أن مستوى تعليم ذوي التلاميذ ومستوى دخلهم له تأثير كبير على ما يحققون من إنجاز، بغض النظر عن المدارس التي يرتادونها. وفي هذه الحالة تعد الإحصائية الوحيدة التي نفتقر إليها هي الوحيدة التي نحتاجها للوصول إلى غايتنا، ألا وهي: ما مدى أداء هؤلاء التلاميذ – جيدًا كان أو سيئًا – والذي يمكن أن يعزى إلى ما يحدث داخل المدرسة (أو داخل قاعة دراسية معينة)؟

فالتلاميذ القادمون من مجتمعات ثرية ذات مستوًى تعليمي جيد سوف يُبلون بلاءً حسنًا في الدراسة منذ اللحظة الأولى التي يوصلهم فيها ذووهم إلى باب روضة الأطفال. لكن الجانب الآخر صحيح كذلك. فهناك مدارس تضم تلاميذ من الطبقات المحرومة يؤدي فيها المدرسون أداءً عظيمًا، لكن نتائج التلاميذ في الاختبارات تظل سيئة – لكنها ليست بنفس السوء الذي كانت ستصير إليه لولا مجهودات المدرسين العظيمة. ما نحتاجه هو إيجاد مقياس لـ «القيمة المضافة» على مستوى المدارس أو حتى على مستوى قاعات الدرس. إننا لا نريد أن نعرف المستوى المطلق لما حققه التلميذ، وإنما نريد أن نعرف كم من هذا الإنجاز الذي تَحقَّق قد أثَّرت عليه العوامل التعليمية التي نحاول تقييمها.

قد تبدو تلك مهمة يسيرة، للوهلة الأولى؛ حيث إننا نستطيع بكل بساطة أن نعطي للتلاميذ اختبارات سابقة أو أخرى لاحقة. فإذا ما كنا نعرف نتائج اختبارات التلاميذ في بداية التحاقهم بمدرسة معينة أو بقاعة دراسية معينة، ثم نقيس أداءهم في نهاية فترة التحاقهم، وهو ما يمكننا من أن نعزو الاختلاف إلى ما قد حدث في المدرسة أو قاعة الدرس.

لكن يا للخسارة! فمرة أخرى هذا الكلام ليس صحيحًا. فالتلاميذ ذوو القدرات المختلفة والقادمون من خلفيات متنوعة يحصِّلون العلم بمعدلات مختلفة. فبعضهم قادرون على فهم المواد التي يدرسونها بوتيرة أسرع من رفاقهم لأسباب لا تتعلق بجودة التدريس. فإذا ما بدأ التلاميذ في المدرسة الثرية أ والمدرسة الفقيرة ب في دراسة علم الجبر في نفس الوقت وعلى نفس المستوى، ثم وجدنا تلاميذ المدرسة أ قد أحرزوا نتائج فاقت

تلاميذ المدرسة ب بعد عام واحد، فقد يكون مردُّ هذا إلى أن مدرسيهم أفضل، أو قد يكون السبب هو قدرتهم على التعلم بشكل أسرع، أو قد يتمثل السبب في الأمرين معًا. يعمل الباحثون الآن على وضع تقنيات إحصائية تقيس الجودة التعليمية بطرق تضع في الاعتبار قدرات التلاميذ وخلفياتهم المتنوعة. لكن إلى أن يتم هذا ستظل محاولاتنا لتحديد «أفضل» المدارس مضللة للغاية.

في فصل الخريف من كل عام تقوم كثير من الصحف والمجلات الصادرة في شيكاجو بنشر قائمة «لأفضل» المدارس الثانوية في المنطقة، وعادة ما يكون هذا الترتيب قائمًا على نتائج الاختبارات. هنا يتبدَّى الجزء المضحك من الأمر؛ من وجهة النظر الإحصائية، فالعديد من المدارس الثانوية، التي عادةً ما تحتل المراكز الخمس الأولى من القائمة، هي مدارس تنتهج منهج الالتحاق الانتقائي؛ أي إنه يتعين على التلاميذ أن يتقدموا بطلبات للالتحاق بها ولا يُقبل منهم إلا العدد القليل. وواحد من أهم معايير الالتحاق هو نتائج الاختبارات الموحدة. دعنا نلخص الأمر فيما يلي: (١) تصنَّف هذه المدارس كمدارس «ممتازة» لأنها تضم تلاميذ ذوي نتائج عالية في الاختبارات. (٢) لا بد للتلميذ كي يلتحق بهذه المدارس من أن يحرز نتائج جيدة في الاختبارات. يشبه هذا الأمر في منطقه أن يُمنح فريقٌ ما لكرة السلة جائزةً لأنه يحسن إنتاج التلاميذ الطوال القامة.

حتى لو كنت تمتلك مؤشرًا واضحًا لما تحاول أن تقيسه أو تديره، فإن التحديات الماثلة أمامك لم تنتهِ بعد. لكن الخبر الجيد هنا هو أن «الإدارة من خلال الإحصاء» يمكن أن تغيِّر السلوك الذي يسلكه الشخص أو تسلكه المؤسسة المقصودان بالإدارة. فإذا ما استطعت أن تقيس قدر المنتجات المعيبة التي تخرج من خط التجميع، وإذا ما كانت تلك العيوب ناتجة عن أمور تحدث في المصنع، فإنك إذا استطعت أن تقدم مكافأة أو علاوة إلى بعض العمال الذين تتمثل مهمتهم في تقليل تلك العيوب، حينها ستغير من السلوك بشكل صحيح. كلنا نستجيب للتحفيز (حتى لو كان في صورة مديح أو مساحة مميزة لركن سيارة). يقيس الإحصاء النتائج التي تهم، ويمنحنا التحفيز أسبابًا لتحسين هذه النتائج.

أو قد يعطينا في بعض الأحيان أسبابًا تجعل الإحصائيات تبدو أفضل. وهذا هو النبأ السيء في المسألة.

إذا ما تم تقييم مديري المدارس — وربما تحديد رواتبهم — على أساس معدل تخرُّج التلاميذ من مدرسةٍ ما، حينها سوف يُولُون كل اهتمامهم لتخريج أكبر عدد من التلاميذ. بالطبع إنهم كذلك يهتمون بتكريس بعض مجهوداتهم لتحسين معدل التخرج، وهو ليس بالضرورة الأمر نفسه. فمثلًا يمكن أن يصنف التلاميذ الذين يتركون المدرسة الثانوية قبل التخرج كتلاميذ «قد انتقلوا لمجال آخر» وليسوا قد تسربوا من التعليم. وهذا ليس بمثال افتراضي، وإنما هو اتهام قد وُجِّه إلى رود بيج وزير التعليم السابق خلال فترة عمله

كمشرف على مدارس هيوستن. كان الرئيس جورج دبليو. بوش قد عَيَّن بيج وزيرًا للتعليم بسبب النجاح الكبير الذي أحرزه في خفض معدلات التسرب من التعليم وزيادة معدلات النتائج الجيدة في الاختبارات.

إذا ما كنتَ تتعقب تلك المقولات المأثورة التي ما أنفكُّ أنا أُلقيها في طريقك، فإليك مقولة أخرى: «عندما تجد فريق عمل برنامج ٦٠ دقيقة لدى باب منزلك، فلن يكون يومك هذا طيبًا». قام دان راذر وفريق الوحدة الثانية لبرنامج ٦٠ دقيقة برحلة إلى هيوستن، وهناك وجدوا أن التلاعب الذي حدث بالإحصائيات أكثر إثارة للإعجاب من التطور التعليمي نفسه[٨]. فقد كانت المدارس الثانوية تقوم بشكل روتيني بتصنيف التلاميذ الذين يهجرون المدرسة كتلاميذ تحولوا إلى مدارس أخرى عائدين إلى بلداتهم الأم أو التحقوا بدبلومة المعادلة العامة (GED)، ولم تسجل الإحصائيات أيًّا من هذه الحالات على أنها تسرُّب من التعليم. ذلك العام الذي تم فيه تفحُّص إحصائيات هيوستن، أعلنت المدينة عن أن معدل التسرب من التعليم فيها يقدر بـ ١,٥ بالمائة، لكن برنامج ٦٠ دقيقة حسب معدل التسرب الحقيقي فوجده يتراوح ما بين الـ ٢٥ والـ ٥٠ بالمائة.

وكذلك كانت المغالطة في معدل نتائج اختبارات التلاميذ، لا تقل عن إبهارًا عن سابقتها. واحدة من الوسائل لتحسين نتائج الاختبارات (في هيوستن وفي كل مكان آخر) هي بتحسين جودة التعليم؛ حتى يتعلم التلاميذ بشكل أفضل، ومن ثم يؤدوا بشكل أفضل في الاختبارات. وذلك أمر طيب، لكن هناك طريقة أخرى (أقل نزاهة) لتحسين نتائج الاختبارات، تتمثل في منع التلاميذ الأضعف مستوى من تأدية الاختبار. فإذا ما أزيلت نتائج التلاميذ الأضعف، فحينها سيرتفع متوسط نتائج الاختبارات للمدرسة أو للمقاطعة بأسرها، حتى لو لم يحقق بقية التلاميذ أي تقدم. في ولاية تكساس يعطى اختبار الإنجاز الدراسي على مستوى الولاية لتلاميذ الصف العاشر. وفي هيوستن تكشَّفت أدلة تشير إلى أن المدارس تحاول ألا يصل التلاميذ الأضعف إلى الصف العاشر. بل إن أحد الأمثلة الفظيعة التي حدثت كان لتلميذ قضى ثلاث سنوات في الصف التاسع ثم تم تصعيده مباشرة للصف الحادي عشر، وهي حيلة شيطانية ماكرة تهدف إلى منع ذلك التلميذ الضعيف المستوى من الخضوع لاختبار الصف العاشر الحاسم دون إجباره على التسرب من التعليم (والذي كان سيظهر في إحصائية مختلفة).

لم يتأكد ضلوع رود بيج في هذه الخديعة الإحصائية خلال فترة إشرافه على هيوستن، لكن الأكيد أنه قام بوضع برنامج صارم يمنح مديري المدارس، الذين يحققون النتائج المرجوة من حيث معدلات التسرب من التعليم ومعدلات نتائج الاختبارات، حوافز مالية، ويقيل أو يخفض من درجة المديرين الذين يخفقون في تحقيقها. لقد استجاب المديرون لتلك المحفزات بكل تأكيد، وهذا هو الدرس الأهم هنا. لكن عليك أن تكون متأكدًا جدًّا أن

أولئك الرفاق الذين يتم تقييمهم لا يستطيعون أن يحسِّنوا من صورتهم (من الناحية الإحصائية) بطرق لا تتناسق مع الأهداف المتاحة لديهم.

ولقد أدركت ولاية نيويورك هذه الحقيقة بالطريقة القاسية، فلقد انتهجت طريقة «بطاقات النقاط» لتقييم معدلات الوفيات لدى مرضى القلب الذين يخضعون لعمليات رأب الشريان التاجي، والتي تعتبر علاجًا شائعًا لمرض القلب[٩] قد يبدو هذا استخدامًا منطقيًّا ومفيدًا للإحصائيات الوصفية. إن عدد مرضى القلب الذين يُتوفَّوْن خلال العمليات الجراحية هو أمر مهم يجب أن نعرفه، ومن المنطقي أن تقوم الحكومة بجمع تلك البيانات ونشْرها، لأنها لو لم تفعل لما استطاع الفرد العادي الوصول إليها. إذن هل هي سياسة جيدة؟ نعم، لولا أنه قد صار من المحتمل أنها تقتل الناس.

من الجلي أن أطباء القلب يهتمون بـ «بطاقات نقاطهم». لكن أسهل الطرق التي يستطيع بها الجرَّاح أن يحسِّن من معدل وفياته ليست بقتل عدد أقل من الناس، على افتراض أن معظم الأطباء يبذلون قصارى جهدهم لأن يُبقوا على حياة مرضاهم. وإنما أسهل الطرق التي يخفِّض بها الطبيب معدل وفيات مرضاه هي رفض إجراء عمليات جراحية للمرضى ذوي الحالات الحرجة. أظهرَ استقصاءٌ أجْرَتْه كلية الطب وطب الأسنان بجامعة روتشستر، أن بطاقات النقاط تلك التي تخدم مصلحة المريض ظاهريًّا يمكن أن تتسبب في أذيته؛ فقد ذكر ما نسبته ٨٣ بالمائة من أطباء القلب الذين شاركوا في الاستقصاء أنه، بسبب إحصائيات معدل الوفيات التي تذاع على الملأ، لا يتلقى بعض المرضى الذين قد يستفيدوا من جراحة رأب الشريان التاجي هذه العملية، وذكر ما نسبته ٧٩ بالمائة من الأطباء أن القرارات الطبية التي يتخذونها بشكل شخصي قد تأثرت بحقيقة أن معدلات الوفيات صارت تُجمع وتُنشر على العلن. يكمن التناقض المؤسف في مسألة هذه الإحصائية الوصفية، التي تبدو نافعة من الناحية الظاهرية، في كون أطباء القلب يستجيبون لها بشكل عقلاني يجعلهم يمنعون الرعاية الصحية عن المرضى الذين هم في أشد الحاجة إليها.

إن أي قائمة إحصائية لها نفس العيوب المحتملة لأي إحصائية وصفية – علاوة على الانحرافات التي تنتج عن الجمع بين مؤشرات متعددة وتحويلها إلى رقم مفرد. إن أي قائمة، كما يقتضي تعريفها، تتسم بالحساسية تجاه الكيفية التي أنشئت بها، فهي تتأثر بالمقاييس التي تتخذها وما تتمتع به تلك المقاييس من ثقل. فمثلًا، لماذا لا نجد في إحصائية تمرير الكرات في البطولة الوطنية لكرة القدم أيَّ مقياس لإكمال اللعبة الثالثة في المباريات؟ وبالنسبة إلى قائمة إحصائيات التنمية البشرية، كيف هو ثقل معدل إجادة القراءة والكتابة بالنسبة إلى نصيب الفرد من الدخل؟ وفي النهاية يبقى السؤال المهم هنا عما إذا كانت البساطة وسهولة الاستخدام، التي تتمثل في اختزال مؤشرات عديدة إلى رقم مفرد، تطغى على ما في تلك العملية من عدم دقة نابعة من داخلها. أحيانًا تكون الإجابة

بالنفي، وهو ما يعيدنا مرة أخرى (كما وعدنا من قبل) إلى ترتيب الجامعات الذي وضعه تقرير شبكة يو إس نيوز آند ورلد (USNWR).

يَستخدم هذا التقرير ستة عشر مؤشرًا، رصد من خلاله نتائج الكليات والجامعات الأمريكية والمدارس المتخصصة، ورتبها وفقًا لتلك المؤشرات. ففي عام ٢٠١٠، على سبيل المثال، استخدمت في هذا الترتيب للجامعات الوطنية وكليات الفن الحر، «الانتقائية الطلابية» ممثلة ١٥ بالمائة من القائمة، والانتقائية الطلابية بدورها قد تم حسابها اعتمادًا على معدل القبول في المدارس ونسبة التحاق الطلاب الذين كانوا ضمن نسبة الـ ١٠ بالمائة الأوائل على صفوفهم الدراسية، ومتوسط نتائج اختبارات SAT وACT للطلبة الملتحقين بالجامعات. تتمثل فوائد هذا التقرير في كونه يوفر الكثير من المعلومات عن آلاف المدارس بشكل بسيط يسهُل الوصول إليه، لدرجة أنه حتى منتقدو التقرير يُقرُّون بقيمة المعلومات التي جُمعت عن جامعات أمريكا وكلياتها. فيجدر بالطلبة القادمين أن يعرفوا معدلات التخرج ومتوسط حجم الصف الدراسي في المؤسسات التعليمية التي يُزمِعون الالتحاق بها.

بالطبع تختلف عملية تقديم معلومات لها معنًى عن اختزال كل تلك المعلومات في تراتب مفرد ذي مصداقية مزعومة. يرى منتقدو هذه التراتبات أنها توضع بشكل يتسم بالفوضى ويشوبها التضليل، وأنها تضر بمصلحة الطلبة على المدى الطويل. يقول مايكل ماكفيرسون الرئيس السابق لكلية ماكاليستر بمنيسوتا: «مما يبعث على القلق في هذه القائمة هو ادعاؤها قدرتَها على ترتيب المؤسسات التعليمية بشكل رقمي، وهو مستوًى من الدقة يتجاوز قدرة تلك البيانات»[١٠]. فلماذا يمثِّل عدد المتخرجين من المدرسة ما نسبته ٥ بالمائة من ترتيبها؟ وإذا كان عدد المتخرجين مهمًّا، فلماذا لم تجعل له نسبة ١٠ بالمائة؟

يَذكر تقرير شبكة يو إس نيوز آند ورلد أن «كل مؤشر من تلك المؤشرات يعيَّن له وزن (ويعبر عن ذلك الوزن عن طريق النسب المئوية)، اعتمادًا على حكمنا على أيُّها أكثر أهمية»[١١] لكن الحكم يختلف تمامًا عن الفوضوية. ويُعدُّ أكثر المتغيرات ثقلًا في ترتيب الجامعات والكليات الوطنية هو «السمعة الأكاديمية». وهذه السمعة تُقرَّر على أساس «استقصاءات تقييمية من أقران» يَملأ بياناتِها إداريُّون في كليات وجامعات أخرى، وتُقرَّر كذلك بناءً على استقصاءات يَملأ بياناتِها مستشارو التوجيه في المدارس الثانوية. وفي نقده العام لتلك القوائم الترتيبية، وجَّه مالكولم جلادويل اتهامًا حادًّا (وإن كان فكاهيًّا) لمنهج تقييم الأقران ذلك. فقد استَشهد باستبيان أرسله أحد رؤساء القضاة السابقين في محكمة ميتشيجن العليا لنحو مائة محامٍ، سائلًا إياهم أن يقوموا بترتيب عشرٍ من كلياتٍ للقانون من حيث الجودة. كانت كلية بن ستايت واحدة من الكليات التي أدرجها المحامون في قوائمهم ووضعوها في موقع يقترب من منتصف القوائم. لكن في ذلك الوقت لم يكن في جامعة بن ستايت كلية قانون من الأساس[١٢]

في كل البيانات التي جمعها تقرير USNWR، لا يبدو أن هذا الترتيب قد قام بقياس ما يهتم به الطلبة المستقبليون؛ مثل ما هو قدر التعليم الذي يتم في كل واحدة من تلك المؤسسات التعليمية؟ قد يعترض مشجعو كرة القادم على قائمة إحصائيات اللاعبين، لكن لا يستطيع أحد أن ينكر أن الأجزاء المكوِّنة لتلك القائمة – إكمال الألعاب وعدد الياردات التي قُطعت والأهداف والإعاقات – هي جزء مهم جدًّا في تقييم إجمالي أداء اللاعبين. لكن هذا الحال لا ينطبق بالضرورة على المعايير التي ينتهجها تقرير الـ USNWR؛ إذ يركز معظم تلك المعايير على المدخلات (على غرار كم من الطلاب قُبلوا، وكم يتقاضى المعلمون، والنسبة المئوية للمعلمين الذين يعملون بدوام كامل) أكثر من المخرجات التعليمية. هناك استثناءان مهمان لما سبق؛ ألا وهما استمرار الملتحقين الجدد بالمدرسة، ومعدل التخرج، لكن حتى هذان المؤشران لا يقيسان عملية التعلم. يقول مايكل ماكفيرسون موضحًا: «إننا لا نعرف من تقرير يو إس نيوز هذا إذا ما كان التعليم الذي يتلقونه خلال تلك السنوات الأربع قد طوَّر من مهاراتهم وأثرى معارفهم».

كل هذا يعتبر جهدًا لا ضير منه سوى أنه يبدو مشجعًا على سلوك لا يصب بالضرورة في مصلحة الطلبة ولا التعليم العالي. فواحدة من الإحصائيات المستخدمة، على سبيل المثال، في حساب الترتيب هي الموارد المالية لكل واحد من الطلبة. تكمن المشكلة هنا في عدم وجود أي مقياس دالٍّ على كم من تلك النقود قد أنفقت على الوجه الصحيح. فالمؤسسات التي تنفق أموالًا أقل بالشكل الأمثل (وهو ما يجعلها تفرض رسومًا أقل على الطلبة) تعاقبها عملية الترتيب تلك. كما أن الكليات والجامعات تجد نفسها مدفوعة لأن تشجع أعدادًا كبيرة من الطلبة على التقدم للالتحاق بها، ومن بينهم طلبة ليس لهم أمل واقعي في القبول لأن هذا يجعل المدرسة تبدو أكثر انتقائية. وهذا يعتبر إهدارًا لموارد مدارس تروِّج لاستمارات طلب التحاق زائفة، ولطلبة يتقدمون للمدارس دونما أية فرصة حقيقية للقبول.

وبما أننا بصدد الانتقال إلى فصل يتناول الاحتمالات، فإنني أراهن على أن تقرير يو إس نيوز آند ورلد الترتيبي سيظل معنا لوقت طويل. وكما يقول ليون بوتستاين رئيس كلية بارد: «إن الناس يحبون الإجابات السهلة. أيُّ الأماكن أفضل؟ رقم ١»[١٢]

الدرس العام المستفاد من هذا الفصل يتلخص في كون هذا الخلل الإحصائي ليس له دخل بالأخطاء الإحصائية. بل إن الحسابات المتقنة يمكن أن تخفي وراءها دوافع مشينة. وكونك قد حسبت المتوسط بدقة لا يلغي حقيقة أن الوسيط هو مؤشر أكثر تحديدًا. كما أن الحكم الشخصي والنزاهة قد اتضح أنهما مهمان بشكل مدهش. والمعرفة المفصلة بالإحصائيات لا تمنع ما فيها من أخطاء تمامًا كما لا تردع المعرفةُ بالقوانين السلوكياتِ الإجرامية. ففي الإحصائيات والجريمة كلتيهما دائمًا ما يدرك الشرير ما يفعل بالضبط.

الفصل الرابع

الارتباط

كيف تعرف نتفليكس ما أحب من الأفلام؟

تصر منصة نتفليكس الرقمية على أنني سوف أعجب بفيلم بوتو، وهو فيلم وثائقي يقدم نظرة متعمقة - ومؤلمة أحيانًا - على حياة رئيسة الوزراء الباكستانية السابقة بنظير بوتو ووفاتها المأساوية. سوف يروقني هذا الفيلم على الأرجح. (لقد أضفته بالفعل إلى قائمة مشاهداتي). فقد كانت الأفلام التي رشحتها لي نتفليكس في السابق رائعة. وحتى الأفلام التي ترشحها لي، والتي شاهدتها سابقًا، دائمًا ما تكون أفلامًا قد استمتعت بمشاهدتها.

كيف تفعلها نتفليكس؟ هل لديها فريق ضخم من المتدربين الذين يجلسون في قبو مقر المنصة الرقمية ويستعينون بموقع جوجل مع إجراء مقابلات مع أهلي وأصدقائي كي يقرروا أنني غالبًا سوف أحب فيلمًا وثائقيًّا عن رئيسة وزراء باكستانية سابقة؟ بالطبع لا. فلم تفعل نتفليكس سوى أنها أتقنت إجراء بعض الإحصائيات المعقدة. نتفليكس لا تعرفني، لكنها تعرف الأفلام التي أعجبتني سابقًا (لأنني وضعت تقييمًا لها). وباستخدام هذه المعلومات، بجانب تقييمات مشاهدين آخرين مع استخدام حاسوب قوي، تستطيع نتفليكس أن تخرج بتنبؤات صادمة في دقتها عن ذوقي في المشاهدة.

سوف أعود لاحقًا إلى الخوارزمية التي تستخدمها نتفليكس في وضع هذه الاختيارات، لكن النقطة المهمة الآن هي أن كل شيء يعتمد على الارتباط. ترشح لي نتفليكس أفلامًا تشبه الأفلام التي أعجبتني، كما أنها ترشح أيضًا أفلامًا أعطاها مشاهدون آخرون تقييمات عالية مشابهة لتقييماتي. لقد تم ترشيح بوتو لي بسبب تقييم الخمس نجوم الذي منحته لفيلمين وثائقيين هما Enron: The Smartest Guys in the Room وFog of War.

يقيس الارتباطُ الدرجةَ التي بها ترتبط ظاهرتان معًا. فمثلًا هناك ارتباط بين درجة الحرارة خلال فصل الصيف، ومبيعات الآيس كريم. فعندما ترتفع واحدة منهما ترتفع معها الأخرى. يرتبط المتغيران إيجابيًّا إذا ما استلزم التغير في إحداهما تغيرًا في الأخرى في نفس

الاتجاه، كما هو الحال في العلاقة بين الطول والوزن. فالأشخاص الطوال (في المتوسط) ذوي وزن أكبر، ويزن القصار أقل. يغدو الارتباط سلبيًّا إذا ما استلزم التغير الإيجابي في واحد من المتغيرات تغيرًا سلبيًّا في الآخر، كما في العلاقة بين التمارين الرياضية والوزن.

لكن الأمر المخادع في هذا النوع من التلازمات يكمن في كون المشاهدات الإحصائية لا تتماشى كلها مع هذا النمط. فأحيانًا يزن قصار القامة أكثر من الطوال. وأحيانًا نجد من لا يمارسون الرياضة أكثر نحولًا ممن يكثرون ممارستها. لكن مع هذا تظل هناك علاقة ذات معنًى بين الطول والوزن وبين التمرين والوزن.

وإذا ما حاولنا أن نضع مخطط انتشار يمثل أطوال وأوزان عينة عشوائية من الأمريكيين البالغين، فقد نتوقع أن نرى شيئًا كما الشكل التالي:

مخطط انتشار يمثل الطول والوزن



وإذا ما حاولنا وضع مخطط انتشار للارتباط بين ممارسة الرياضة (قياسًا بعدد الدقائق التي تنفق في التمرينات المكثفة أسبوعيًّا) والوزن، فسنتوقع وجود ارتباط سلبي لدى أولئك الذين يمارسون الرياضة أكثر فيفقدون وزنًا. لكن ذلك النمط الذي يمثَّل بنقاط متناثرة فوق صفحة هو أداة غير عملية. (لو حاولت نتفليكس أن تضع لي الترشيحات عن طريق وضع مخططات انتشار لتقييمات آلاف الأفلام التي وضعها ملايين المشاهدين، لانطمر مقرها تحت كل هذا العدد الهائل من المخططات). لكن قوة الارتباط كأداة إحصائية تكمن في قدرته على تلخيص التلازم بين متغيرين في إحصائية وصفية واحدة، ألا وهي: معامل الارتباط.

يتميز معامل الارتباط بخاصيتين رائعتين. تتعلق أولاهما بالحسابات الرياضية التي أخَّرناها إلى ملحق الفصل، إذ يتمثل هذا المعامل في رقم مفرد يتدرج من -١ إلى ١. والارتباط ذو القيمة ١ يوصف دائمًا بأنه ارتباط تام، مما يعني أن كل تغير يطرأ على أحد المتغيرات يستلزم تغيرًا مكافئًا في المتغير الآخر في نفس الاتجاه.

أما الارتباط ذو القيمة -١، فهو ارتباط سلبي تام، وهو ما يعني أن أي تغير يطرأ على واحد من المتغيرات يستلزم تغيرًا مكافئًا في الآخر في الاتجاه المضاد.

كلما اقترب الارتباط من ١ أو -١، زادت قوة الاستلزام. أما الارتباط ذو القيمة (أو ما يدنو منها) فهو يعني عدم وجود أي تلازم بين المتغيرين، كما هو الحال مع مقاس حذاء المرء ونتائجه في اختبار SAT.

أما الخاصية الرائعة الثانية التي يتميز بها معامل الارتباط، فتتمثل في كونها لا تلتزم وحدات معينة. فبإمكاننا أن نحسب الارتباط بين الطول والوزن – حتى والطول يقاس بالبوصات بينما يقاس الوزن بالأرطال. بل إننا قادرون على أن نحسب الارتباط بين عدد أجهزة التليفزيون الموجودة في منازل تلاميذ المدارس الثانوية ونتائجهم في اختبار SAT، وهو الارتباط الذي أؤكد أنه إيجابي. (سأزيدك عن هذا الموضوع قريبًا). هكذا يبدو أن معامل الارتباط يقوم بشيء يشبه الإعجاز، فهو يختزل كمًّا فوضويًّا معقدًا من البيانات التي قيست من خلال وحدات مختلفة (كما في حال مخطط انتشار الطول والوزن) في إحصائية وصفية أنيقة واحدة.

كيف؟

كالعادة وضعتُ أكثر المعادلات شيوعًا لحساب معامل الارتباط في ملحق بنهاية الفصل. ليست تلك بمعادلة ستحسبها يدويًّا. (فبعد أن تُدخل البيانات يقوم أحد برامج الحاسوب الأساسية على غرار Microsoft Excel بحساب الارتباط بين متغيرين). لكن مع هذا ليس استنباطه بهذه الصعوبة، فمعادلة حساب معامل الارتباط تقوم بما يلي:

١- تَحسب المتوسط والانحراف المعياري للمتغيرين كليهما. فإذا ما ظللنا مع مثال الطول والوزن؛ فسنعرف متوسط الطول لأفراد العينة، ومتوسط الوزن لأفراد العينة، والانحراف المعياري للطول والوزن كليهما.

٢- تحوِّل جميع البيانات حتى يتسنى تمثيل المشاهدات الإحصائية من خلال بُعدها (في الانحراف المعياري) عن المتوسط. انتبه معي ولن تجد الأمر معقدًا. لنفترض أن متوسط الطول في العينة يقدر بـ ٦٦ بوصة (مع انحراف معياري يقدر بخمس بوصات)، وأن متوسط الوزن يقدر بـ ١٧٧ رطلًا (مع انحراف معياري يقدر بعشرة أرطال). والآن افترض أن طولك ٧٢ بوصة وأنك تزن ١٦٨ رطلًا. يمكننا أيضًا أن نقول إن طولك هو انحراف معياري بمقدار ١,٢ فوق متوسط الطول

[(٧٢-٦٦)/٥] وهو انحراف معياري بمقدار ٠,٩ تحت متوسط الوزن أو – ٠,٩ طبقًا للمعادلة [(١٦٨ – ١٧٧)/١٠]. صحيح أنه من غير المعتاد أن نجد شخصًا فوق متوسط الطول وتحت متوسط الوزن، لكن لأنك دفعت مبلغًا محترمًا لشراء هذا الكتاب، فقد فكرت أن أقل ما يجب أن أفعله هو أن أجعلك طويلًا نحيلًا. لاحظ أن طولك ووزنك اللذين عبَّرنا عنهما سابقًا بالبوصات والأرطال تحوَّلا إلى ١,٢ و-٠,٩. وذلك هو ما يجعل الوحدات تختفي.

٣- هنا أنفض يدي وأترك الحاسوب يؤدي العمل كله. فالمعادلة تقوم بحساب العلاقة بين الطول والوزن لدى جميع أفراد العينة عن طريق قياسها بالوحدات المعيارية. فإذا كان أفراد العينة طوالًا جدًّا بما قدره، على سبيل الافتراض، ١,٥ أو ٢ انحراف معياري فوق المتوسط، إذن فكم ستكون الأوزان قياسًا بالانحراف المعياري عن متوسط الوزن؟ وإذا اقترب أفراد العينة من متوسط الطول، فكم ستكون أوزانهم مَقيسة بالوحدات المعيارية؟

إذا كانت المسافة من المتوسط لمتغير واحد تميل لأن تكون متوافقة مع مسافة المتغير الآخر عن المتوسط للمتغير الآخر (مثلًا الأفراد الذين يبعدون عن متوسط الطول في أيٍّ من الاتجاهين يميلون أيضًا لأن يكونوا بعيدين عن المتوسط في نفس الاتجاه عن متوسط الوزن) حينها علينا أن نتوقع ارتباطًا إيجابيًّا قويًّا.

إذا ما كان البعد عن المتوسط في أحد المتغيرات يتطابق مع تباعد آخر عن المتوسط في متغير آخر ولكن في الاتجاه المعاكس (كما في حالة كون مَن يعلون كثيرًا عن المتوسط في ممارسة الرياضة يقلون كثيرًا عن المتوسط لدى الوزن) فلنا، حينها، أن نتوقع ارتباطًا سلبيًّا قويًّا.

أما إذا كان المتغيران لا ينحرفان عن المتوسط انحرافًا دالًّا (كما في حالة مقاس الحذاء وممارسة الرياضة)، فلنا أن نتوقع ارتباطًا ضعيفًا أو لا ارتباط على الإطلاق.

لقد قاسيت كثيرًا في هذا الجزء من الفصل، لذا سوف نعود قريبًا إلى استئجار الأفلام. لكن قبل أن نرجع إلى نتفليكس، دعنا نتناول وجهًا آخر من وجوه الحياة يحظى فيه الارتباط بقدر كبير من الأهمية، ألا وهو اختبارات SAT. نعم اختبارات SAT. اختبارات SAT للمنطق الذي عُرف سابقًا باسم اختبار القابلية الدراسية هو اختبار موحد مقسم إلى ثلاثة أقسام، وهي الرياضيات والقراءة والكتابة. لعلك قد خضعت لهذا الاختبار أو أنت بصدد الخضوع إليه قريبًا. لكنك لم تتفكر كثيرًا، على الأرجح، في السبب الذي جعلك تخضع له. غرض هذا الاختبار هو قياس قدراتك الأكاديمية والتنبؤ بأدائك في الجامعة. قد يتسائل بعضهم تساؤلًا منطقيًّا (خاصة أولئك الكارهين للاختبارات الموحدة) قائلين: ألم تُجعل المدارس الثانوية لهذا الغرض؟ ولماذا نعلق أهمية كبيرة على اختبار

مدته أربع ساعات، في حين يتاح للمختصين بالقبول في الجامعات سِجِل لتقديرات أحرزها الطالب في أربع سنوات؟

في الفصلين الأول والثاني من هذا الكتاب نجد الإجابة عن هذه الأسئلة، والتي تتمثل في كون التقديرات الدراسية تعتبر إحصائية وصفية يشوبها عوار. فالطالب الذي يحصل على تقديرات متواضعة في ظل جدول دراسي صعب به مواد الرياضيات والعلوم، قد تكون قدراته وما يتوقع منه من الناحية الأكاديمية يفوقان ما يماثلهما لدى طالب آخر في نفس المدرسة حاز تقديرات أفضل في مواد دراسية أقل صعوبة. هكذا يتضح ما في المدارس من تناقضات واسعة. يقول مجلس الجامعات، الذي هو الهيئة المسئولة عن وضع اختبارات SAT وإدارتها، إن هذا الاختبار قد أُوجِدَ بهدف «جعل عملية الالتحاق بالجامعات عملية ديمقراطية لجميع الطلبة». هو أمر من قبيل العدل. فهذا الاختبار يتيح مقياسًا معياريًّا للقدرات التي يمكن مقارنتها لدى جميع الطلبة المتقدمين للجامعة. لكن أهو مقياس قدرات جيد؟ فإذا ما أردنا مقياسًا قابلًا للمقارنة بسهولة لدى جميع الطلبة، نستطيع أن نجعل جميع طلبة السنة النهائية في المدارس الثانوية يشتركون في سباق عدْو لمائة ياردة، وهو أمر أيسر وأزهد تكلفة من إدارة اختبار SAT. لكن المشكلة تكمن بالطبع في أن الأداء في سباق المائة ياردة ذاك غير ذي ارتباط بالأداء الجامعي. فالحصول على البيانات هو أمر يسير، لكنها لن تمنحنا شيئًا ذا معنًى.

إذن إلى أي حد نجحت اختبارات SAT في هذا الصدد؟ للأسف سوف تُبلي تلك الاختبارات بلاءً حسنًا في التنبؤ بالتقديرات الدراسية لطلبة السنة الجامعية الأولى من الأجيال التي ستأتي في المستقبل. فمجلس الجامعات يقوم بإعلان الارتباطات المتعلقة بهذا الأمر بمقياس يتدرج من (حيث لا ارتباط على الإطلاق) إلى ١ (وهو الارتباط التام)، فالارتباط بين متوسط تقديرات المدرسة الثانوية ومتوسط تقديرات أولى سنوات الجامعة هو ٠,٥٦، (وكي نضع هذا الأمر في منظوره نقول إن الارتباط بين الطول والوزن لدى الذكور البالغين في الولايات المتحدة يقدر بنحو ٠,٤). وكذلك فالارتباط بين مركبات اختبار SAT (القراءة النقدية والرياضيات والكتابة) ونقاط التقييم العام للسنة الجامعية الأولى تقدر أيضًا بـ ٠,٥٦[١٠]. قد يبدو ما سبق مرجحًا لنبذ اختبارات SAT بما أن الاختبار لا يفعل خيرًا مما تفعله التقديرات المدرسية في المرحلة الثانوية في التنبؤ بالأداء المنتظر في المرحلة الجامعية. والواقع أن أفضل التنبؤات يخرج منها مزيج بين الـ SAT والتقديرات المدرسية في المرحلة الثانوية، والذي يبلغ ارتباطه بالأداء الدراسي في السنة الجامعية الأولى ٠,٦٤، وإني آسف لذلك.

واحدة من النقاط المحورية في النقاش العام حول المسألة هو أن الارتباط لا ينطوي على السببية، فالتلازم بين متغيرين سواء كان إيجابيًّا أو سلبيًّا لا يعني بالضرورة أن التغير

في أحدهما يسبب تغيرًا في الآخر. فإنني مثلًا قد ألمحت سابقًا إلى ترجيح وجود ارتباط إيجابي بين نتائج التلاميذ في اختبار SAT وعدد أجهزة التلفاز التي تملكها عوائلهم. لا يعني هذا أن الوالدان المبالغان في الحرص على مستقبل أبنائهم سوف يبتاعان خمسة أجهزة تلفاز إضافية طلبًا لتحسين نتائج أبنائهما في الاختبار، ولا يعني كذلك أن مشاهدة التلفاز بكثرة ستزيد من الإنجازات الأكاديمية.

فالتفسير الأكثر منطقية لمثل هذا الارتباط هو أن الوالدين الأكثر تعليمًا، والقادرين على امتلاك المزيد من أجهزة التلفاز، يرجَّح أن يحقق أبناؤهما نتائج أفضل في الاختبار. والمتغيران كلاهما، التلفاز ونتائج الاختبار، تَسَبَّب فيهما على الأرجح متغير ثالث، ألا وهو مستوى تعليم الوالدين. لا أستطيع أن أبرهن على الارتباط بين وجود أجهزة التلفاز في المنزل ونتائج اختبار SAT. (فمجلس الجامعات لا يستطيع أن يقدم لي هذه البيانات). لكنني مع ذلك أستطيع أن أبرهن على أن الطلبة الذين ينتمون لعائلات ثرية يحققون نتائج أفضل من أقرانهم الأقل ثراءً. أشار مجلس الجامعات الأمريكي إلى أن الطلبة المنتمين إلى عائلات ذات دخل يتجاوز الـ ٢٠٠٠٠٠ دولار يبلغ متوسط إنجازهم في قسم الرياضيات باختبار SAT ٥٨٦، في مقابل ٤٦٠ الذي هو متوسط إنجاز أقرانهم في نفس القسم من نفس الاختبار والذين يبلغ دخل عائلاتهم ٢٠٠٠٠ دولار أو أقل[٢] وفي الوقت نفسه فإنه من المرجح أن تمتلك العائلات التي يتجاوز دخلها الـ ٢٠٠٠٠٠ دولار مزيدًا من أجهزة التلفاز في منازلهم (المتعددة) تفوق عددها لدى العائلات التي يبلغ دخلها ٢٠٠٠٠ دولار أو أقل.

لقد شرعت في كتابة هذا الفصل قبل عدة أيام. وفي أثناء ذلك أتيحت لي فرصة لمشاهدة فيلم بوتو. يا للروعة، إنه فيلم مذهل عن عائلة مذهلة! تمتد سردية ذلك الفيلم المتفرد منذ الانفصال الذي تم بين الهند وباكستان عام ١٩٤٧، حتى اغتيال بنظير بوتو عام ٢٠٠٧. إنه فيلم متميز. وفيه قد أدمج صوت بوتو بحِرفية في هيئة خطابات ومقابلات صحفية. على كل حال قد خلعتُ على الفيلم تقييمًا ذا خمسة نجوم، وهو الأمر الذي تنبأت به نتفليكس.

إن ما تقوم به نتفليكس على أكثر المستويات أساسية هو استغلال مفهوم الارتباط. فإنني أقوم أولًا بتقييم مجموعة من الأفلام. فتشرع نتفليكس حينها في مقارنة تقييماتي بتقييمات عملاء آخرين، كي تحدد من بين أولئك العملاء من ترتبط تقييماتهم بتقييماتي. فأولئك العملاء يميلون لأن يعجبوا بالأفلام التي تعجبني. ما إن تتوصل نتفليكس لهذه المعلومة حتى تبدأ في أن ترشح لي تلك الأفلام التي قيَّمها أولئك العملاء الذين يُشْبه فِكرهم فِكري، والتي لم أشاهدها بعد.

ما سبق هو «الصورة الكبيرة». لكن المنهجية الفعلية أكثر تعقيدًا بكثير. في عام ٢٠٠٦

أقامت نتفليكس مسابقة دعت فيها العامة لتصميم آلية تحسِّن نظام ترشيحات نتفليكس بنسبة ١٠ بالمائة (وهو ما يعني أن يقوم هذا النظام الجديد بالتنبؤ بتقييم المُشاهد للفيلم الذي يراه بدقة تزيد عن النظام السابق بنسبة ١٠ بالمائة). ووضعت للفائز جائزة قدرها ١٠٠٠٠٠٠ دولار.

مُنح كل فرد وكل فريق تَقدَّم للمسابقة «بيانات تدريبية» تتألف من أكثر من ١٠٠ مليون تقييم لثمانية عشر ألف فيلم وضعها ٤٨٠٠٠٠ من عملاء نتفليكس. في نفس الحين «أخفت» المنصة عليهم مجموعة أخرى تتألف من ٢,٨ مليون تقييم، وهو ما يعني أن نتفليكس تعرف كيف قيَّم العملاء تلك الأفلام لكن المتسابقين لم يعرفوا. تم الحكم على المتسابقين طبقًا لمدى صحة تنبؤات خوارزمياتهم بتقييمات المشاهدين لتلك الأفلام التي أُخفيت عنهم. وعلى مدى ثلاث سنوات تقدمت آلاف الفرق من أكثر من ١٨٠ دولة بمقترحاتها. كان هناك متطلبان لازمان للاشتراك في تلك المسابقة. أولهما أن الفائز ملزم بمنح نتفليكس رخصة استخدام خوارزميته. وثانيهما أن الفائز يتعين عليه أن «يشرح للعالم كله كيف فعلها والسبب في نجاحها»[٣]

وفي عام ٢٠٠٩ أعلنت نتفليكس اسم الفائز وكان فريقًا مؤلفًا من سبعة من الإحصائيين وعلماء الحاسوب من الولايات المتحدة والنمسا وكندا وإسرائيل. للأسف لا أستطيع أن أشرح ذاك النظام الفائز ولا حتى في الملحق؛ فالدراسة التي تشرح ذلك النظام تبلغ اثنتين وتسعين صفحة طولًا*. تبهرني جودة ترشيحات نتفليكس، لكن هذا النظام ما هو إلا تنوع منمق لما ظل الناس يفعلونه منذ فجر صناعة الأفلام، ألا وهو البحثُ عن شخص يماثل ذوقُه ذوقَك ثم طلبُ الترشيحات منه. إنك تميل لأن يروقك ما يروقني وتكره ما أكره، فما رأيك إذن بفيلم جورج كلوني الجديد؟

ذاك هو جوهر الارتباط.

ملحق الفصل الرابع

كي تحسب معامل الارتباط بين مجموعتين من الأرقام، عليك أن تقوم بالخطوات التالية، وكل منها يشرحها استخدام بيانات الطول والوزن لخمسة عشر تلميذًا افتراضيًّا واردة في الجدول التالي.

١- قم بتحويل طول كل واحد من التلاميذ إلى وحدات معيارية: (الطول – المتوسط)/ الانحراف المعياري.

* يمكنك أن تقرأها من موقع: http://www.netflixprize.com/asserts/Grandprize2009_BPC_PragmaticTheory.pdf

٢- قم بتحويل وزن كل واحد من التلاميذ إلى وحدات معيارية: (الوزن – المتوسط)/ الانحراف المعياري.

٣- قم بحساب الناتج من ضرب (الوزن بالوحدات المعيارية) × (الطول بالوحدات المعيارية) لكل تلميذ. سوف ترى حينها كيف أن هذا الرقم سوف يبدو الأضخم ممثلًا في قيمة مطلقة عندما يكون طول التلميذ ووزنه بعيدين نسبيًّا عن المتوسط.

٤- معامل الارتباط هو مجموع النواتج السابقة مقسومًا على عدد المشاهدات (١٥ في هذه الحالة).

يبلغ الارتباط بين الطول والوزن لهذه المجموعة من التلاميذ ٨٣. ولأن معامل الارتباط يتدرج من -١ إلى ١، فإن ذلك الارتباط السابق يتضح أنه ارتباط إيجابي ذو درجة عالية كما هو متوقع.

أ	ب	ج	د	هـ	و
التلميذ	الطول	الوزن	الطول بوحدات معيارية	الوزن بوحدات معيارية	(الوزن بوحدات معيارية) × (الطول بوحدات معيارية)
نيك	74	193	1,21	0,99	1,19
إلانا	66	133	-0,63	-0,67	0,42
دينا	68	155	-0,17	-0,06	0,01
ريبكا	69	147	0,06	-0,29	-0,02
بن	73	175	0,98	0,49	0,48
تشارو	70	128	0,29	-0,81	-0,24
سحر	60	100	-2,00	-1,59	3,18
ماجي	63	128	-1,32	-0,81	1.07
فيصل	67	170	-0,40	0,35	-0,14
تيد	70	182	0,29	0,68	0,20
نارسيسو	70	178	0,29	0,57	0,17
كاترينا	70	118	0,29	-1,09	-0,32

أ	ب	ج	د	هـ	و
التلميذ	الطول	الوزن	الطول بوحدات معيارية	الوزن بوحدات معيارية	(الوزن بوحدات معيارية) × (الطول بوحدات معيارية)
سي جيه	75	227	1,44	1,93	2,77
صوفيا	62	115	1,54-	1,17-	1,81
ويل	74	211	1,21	1,49	1,80
المتوسط	68,73	157,33			الإجمالي = 12,39
الانحراف المعياري	4,36	36,12		معامل الارتباط = الإجمالي/ عدد التلاميذ = 15/12,39 = 0,83	

تتطلب معادلة حساب معامل الارتباط انحرافًا بسيطًا عن الموضوع يتخلله بعض الرموز. فالرمز Σ، والمعروف برمز المجموع، مهم جدًّا في الإحصائيات. فهو يمثل مجموع الكميات التي تليه. فمثلًا إذا كانت هناك مجموعة من المشاهدات $\chi_1, \chi_2, \chi_3, \chi_4$ وهكذا، فإن $\Sigma(x_i)$ ينبئنا بأننا نحتاج لأن نجمع المشاهدات الأربعة بهذا الشكل: $\chi_1+\chi_2+\chi_3+\chi_4$، وهكذا فإن $\Sigma(x_i)=\chi_1+\chi_2+\chi_3+\chi_4$. ومعادلتنا لمتوسط مجموعة المشاهدات i يمكن أن تمثل بالشكل التالي: المتوسط = $\Sigma(x_i)/n$

بل إننا نستطيع أن نجعل المعادلة أكثر يسرًا من خلال كتابة $\sum_{i=1}^{n}(x_i)$ وهو الرمز الذي يجمع الكميات $\chi_1+\chi_2+\chi+\cdots\chi_n$ ، أو بعبارة أخرى كل ما يبدأ بـ x_1 (لأن i = 1) حتى x_n (لأن i = n). معادلتنا لمتوسط مجموعة من المشاهدات يمكن أن يمثل كما يلي:

المتوسط = $\sum_{i=1}^{n}(x)/n$

وبالنظر إلى هذا الترميز العام، فإن معادلة حساب معامل الارتباط r لمتغيرين x و y هي كالتالي:

$$r=\frac{1}{n}\sum_{i=1}^{n}\frac{(x_i-\bar{x})}{\sigma_x}\frac{(y_i-\bar{y})}{\sigma_y}$$

حيث:

n = عدد المشاهدات

$\bar{x}$ هو متوسط المتغير x

y هو متوسط المتغير y

σ_x هو الانحراف المعياري للمتغير x

σ_y هو الانحراف المعياري للمتغير y

أي برنامج كمبيوتر ذي أدوات إحصائية، يمكنه كذلك أن يحسب معامل الارتباط بين متغيرين. ففي مثال طول التلاميذ ووزنهم ستجد برنامج مايكروسوفت إكسل يعطيك نفس النتيجة التي خرج بها الحساب اليدوي لمعامل الارتباط لطول الخمسة عشر تلميذًا ووزنهم ٠,٨٣.

الفصل الخامس

الاحتمالات الأساسية

لا تشترِ شهادة الضمان الممتد لطابعتك ذات التسعة والتسعين دولارًا

في عام ١٩٨١ أنفقت شركة جوزيف شليتز للجعة ١,٧ مليون دولار على تلك الحملة الإعلانية، التي بدت جريئة جدًّا بشكل صادم محفوف بالمخاطرة، لصالح علامتها التجارية شليتز التي كانت آخذة في التراجع. فخلال استراحة ما بين الشوطين في المباراة النهائية للبطولة الوطنية لكرة القدم، وأمام ١٠٠ مليون مشجع حول العالم، قامت الشركة ببث مسابقة تذوُّق لجعتها أمام منافِستها الرئيسية جعة ميشيلوب[١]. بل إنها بالغت في جراءتها ولم تنتقِ متذوقين عشوائيين لترجيح أحد المشروبين، وإنما انتقت ١٠٠ من معتادي شرب جعة ميشيلوب. كانت تلك ذروة تلك الحملة التي امتدت طوال مباريات المربع الذهبي[٢]. تألفت الحملة من خمسة اختبارات تذوُّق مذاعة عبر شاشات التلفاز، وفي كل منها كان مائة من مستهلكي أحد أنواع الجعة المنافِسة لشليتز (بادوايزر وميلر وميشيلوب) يتذوقون النوعين وهم معصوبو الأعين. وكان كل نزال تذوُّقي يروَّج له بغزارة تمامًا كما كان يروَّج لمباريات المربع الذهبي التي كانت تذاع خلالها (على غرار: «شاهد شليتز تقابل بادوايزر على الهواء خلال مباريات المربع الذهبي للبطولة الوطنية»).

كانت الرسالة التسويقية واضحة ومفادها أن شاربي الجعة الذين يظنون أنهم يفضلون نوعًا آخر من الجعة سوف يفضلون جعة شليتز عندما يتذوقونها معصوبي الأعين. بل إن شركة شليتز قامت بالاستعانة بأحد حكام البطولة الوطنية لكرة القدم السابقين كي يشرف على الاختبار الأخير الذي أذيع في المباراة النهائية. هذه المخاطرة، التي انطوى عليها إجراء اختبارات تذوُّق لأناس معصوبي الأعين أمام جمهور ضخم في بث مباشر على التلفاز، قد توحي إلينا بأن شليتز قد أنتجت جعة ذات مذاق طيب لا يضاهى، أليس كذلك؟

ليس بالضرورة. لم تحتج شليتز سوى جعة متوسطة المستوى وفهم ممتاز للإحصاء كي تدرك أن هذه الخدعة – وهو مصلح لا أستعمله مستخِفًا حتى لو كنت أتكلم عن الترويج للجعة – سوف تصب، ولا شك، في مصلحتها. معظم أنواع الجعة التي تندرج في فئة نوع الجعة التي تنتجها شليتز تتشابه في مذاقاتها، والمفارقة أن تلك الحقيقة هي نفسها التي استغلتها الحملة الترويجية. لنفترض أن شارب الجعة العادي لا يستطيع أن يفرق بين شليتز وبادوايزر وميشيلوب وميلر. في هذه الحالة لن يتعدى اختبار التذوق لمعصوبي الأعين أولئك كونه مجرد رمْي لعملة في الهواء. فنصف المتذوقين تقريبًا سيختارون شليتز وسيختار النصف الآخر «منافستها». هذه الحقيقة لا تقيم وحدها حملة إعلانية فعالة. «فلأنك لا تستطيع التفريق بينهما، بإمكانك أن تشرب شليتز». وبالطبع لن تجري شليتز هذا الاختبار على عملائها الأوفياء لها؛ إذ إن نصفهم قد يختارون منافستها. ستجد نفسك في موقف سيء لو اختار عملاؤك الأوفياء منتجًا ينافسك في اختبار تذوق تعصب فيه أعينهم – وهو بالضبط ما كانت شليتز تحاول فعله بمنافسيها.

لكن شليتز فعلت أمرًا آخر أكثر براعة. فقد قام العبقري الذي كان وراء هذه الحملة باختيار المتذوقين فقط من بين من أعلنوا أنهم يفضلون الأنواع المنافسة. فإذا ما كان التذوق العمياني ذاك ما هو إلا رمي لعملة في الهواء فسيختار نصف شاربي بادوايزر أو ميلر أو ميشيلوب جعة شليتز. وهذا سوف يحسِّن من صورة شليتز جدًّا. تصور أن نصف شاربي بادوايزر يفضلون شليتز.

ستتحسن صورتها أفضل عندما يذاع هذا وسط المباراة النهائية ومع حكم سابق (يرتدي زي الحكام) يحكم الاختبار. لكن لا يزال هذا أمرًا مذاعًا على التلفاز. وحتى لو قرر إحصائيو شليتز من خلال كم كبير من تجارب أجروها سابقًا أن شارب جعة ميشيلوب العادي سوف يختار شليتز في نصف مرات التذوق، ماذا لو اتضح أن ١٠٠ من شاربي ميشيلوب الذين يخضعون للاختبار معصوبي الأعين على شاشة التلفاز في الاستراحة بين شوطي المباراة النهائية هم أناس غرباء الأطوار؟ نعم، يعتبر اختبار التذوق العمياني ذاك مكافئًا لرمي العملة، لكن ماذا لو اختار معظم المتذوقين ميشيلوب بمحض الصدفة؟ فمن المعروف أننا لو طلبنا من ١٠٠ شخص أن يقوموا برمي العملة فمن المحتمل جدًّا أن تقع عملات ٨٥ أو ٩٠ منهم على ذات الوجه. هذا النوع من الحظ العاثر سوف يكون بمثابة كارثة لعلامة شليتز التجارية (فضلًا عن كونه إهدارًا لـ ١٫٧ مليون دولار على تغطية تلفزيونية).

هنا يتدخل الإحصاء للإنقاذ. آه لو كان هناك نوع من أنواع الأبطال الخارقين في مجال الإحصاء*، حينها سوف ينطلق أو تنطلق إلى مقر شركة شليتز ويكشف عن تفاصيل

* في ذهني تصوُّر لبطل خارق يسمى «رجل السيجمات الستة». يمثل حرف سيجما σ الإغريقي الانحراف المعياري. ورجل السيجمات الستة يمثل ستة انحرافات معيارية فوق المعدل الإحصائي من ناحية القدرة الإحصائية والقوة والذكاء.

ما يسميه الإحصائيون تجربة ثنائية الحدين (تسمى أيضًا محاولة برنولي). وتعد الخاصية المميزة لتلك التجربة الثنائية الحدين هي أننا لا نملك إلا عددًا ثابتًا من المحاولات (١٠٠ متذوق مثلًا)، وكل محاولة منها لها نتيجتان محتملتان (شليتز أو ميشيلوب)، واحتمالات «النجاح» متساوية في كل محاولة. (هنا أفترض أن احتمالات اختيار مشروب عن الآخر هي ٥٠ بالمائة، وأعرف «النجاح» في هذا السياق باختيار جعة شليتز). كما أننا نفترض أن كل واحدة من تلك المحاولات «مستقلة»؛ بمعنى أن كل واحد من المتذوقين في ذلك الاختبار لا يؤثر اختياره على اختيارات الآخرين.

بهذه المعلومات وحدها يستطيع بطل الإحصاء الخارق أن يحسب احتمالات كل النواتج المختلفة للمحاولات المائة على غرار ٥٢ لصالح شليتز و٤٨ لصالح ميشيلوب، أو ٣١ لشليتز و٦٩ لميشيلوب. أما مَن هم ليسوا بالأبطال الخارقين منا فيستطيعون استخدام الحاسوب لإنجاز نفس المهمة. تُقدَّر فرصة أن يختار المتذوقون المائة كلهم جعة ميشيلوب ١ لكل ١٢٦٧٦٥٠٦٠٠٢٢٨٢٢٩٤٠١٤٩٦٧٠٣٢٠٥٣٧٦ مرة. هناك احتمالات أكبر منها لأن يقتل جميع المتذوقين في استراحة ما بين الشوطين عن طريق كويكب ساقط من السماء. لكن أهم من هذا أن نفس الحسابات الأولية يمكن أن تعطينا احتمالات تراكمية لعدد من المخرجات، كاحتمالات أن ينتقي أربعون متذوقًا أو ما يزيد جعة شليتز. تلك الأرقام قد سكَّنت، ولا شك، من مخاوف مختصِّي التسويق بشركة شليتز.

لنفترض أن مسئولي شليتز كانوا سيسعدون إذا ما اختار أربعون متذوقًا من أصل المائة جعتهم – وهو رقم ممتاز إذا ما افترضنا أن جميع أولئك الرجال الذين خضعوا للاختبار معصوبي الأعين على الهواء مباشرة قد أقروا بأنهم من شاربي ميشيلوب. هذا الناتج الممتاز كان محتملًا جدًّا. لو كان اختبار التذوق رميًا لعملة من الناحية العملية لأنبأتنا الاحتمالات بأن هناك فرصة تبلغ ٩٨ بالمائة أن ٤٠ متذوقًا على الأقل سوف ينتقون شليتز، وفرصة تقدر بـ ٨٦ بالمائة لأن ينتقي ٤٥ رجلًا ذات الجعة*. لم تنطوِ تلك المناورة على أية خطورة من الناحية النظرية.

إذن ماذا حدث مع شليتز؟ خلال استراحة ما بين الشوطين في المباراة النهائية اختار ما نسبته ٥٠ بالمائة من شاربي جعة ميشيلوب معصوبي الأعين جعة شليتز في اختبار التذوق ذاك.

مما سبق نخرج بدرسين مهمين؛ أولهما أن الاحتمالات أداة قوية للغاية، وثانيهما أن الكثير من أنواع الجعة التي شاعت في ثمانينيات القرن العشرين لم يكن أحدها يختلف عن الآخر. سوف يركز هذا الفصل بشكل أساسي على الدرس الأول.

* لأجل كل تلك الحسابات قمت باستخدام حاسبة إلكترونية ثنائية الحدين عبر الإنترنت على موقع: http://stattrek.com/Tables/Binomial.aspx.

الاحتمالات هي دراسة الأحداث والنواتج التي يتخللها عنصر عدم اليقين. فالاستثمار في سوق الأوراق المالية ينطوي على عدم يقين. وكذلك هو رمي العملة في الهواء؛ فهي تستقر إما كاشفة وجهًا عليه رأس وإما كاشفة آخر عليه ذيل. وإذا ما ألقيت العملة أربع مرات متتالية في الهواء، لانطوى هذا على أربع طبقات من عدم اليقين؛ لأن كل مرة منها قد تأتي بواحدة من النتيجتين. فإذا قمت أنت بإلقاء العملة في الهواء أربع مرات متتالية، فلن أعرف عن يقين مسبق كيف ستكون نتيجة كل رمية (ولن تعرفها أنت كذلك). لكنني أستطيع أن أقرر مسبقًا أن بعض النواتج (مرتين كاشفة وجه الرأس ومرتين كاشفة وجه الذيل) أكثر أرجحية من نواتج أخرى (أربع مرات كاشفة وجه الرأس). ولقد أدرك الرفاق في شليتز هذه الحقيقة التي مفادها أن تلك الأفكار القائمة على الاحتمالات ذات نفع كبير. والواقع أنك لو فهمت أن السبب في كون احتمالات سقوط العملة أربع مرات متتالية كاشفة وجه الرأس – إذا ما كانت العملة سليمة غير معطوبة – هي ١ إلى ١٦، حينها تستطيع (ببعض الجهد) أن تفهم كل شيء من كيفية عمل صناعة التأمين، إلى إذا ما كان من الأفضل لفرق كرة القدم أن يركلوا الركلة الإضافية لكسب نقطة واحدة عقب كل هدف يحرزوه، أم أن عليهم أن يلعبوا اللعبة التي قد تكسبهم نقطتين.

دعنا نبدأ بالجزء السهل، والذي ينص على أن لجميع الأحداث احتمالات معروفة. فاحتمالات ظهور وجه الرأس عند رمي العملة تقدر بـ ½، وتقدر احتمالات ظهور رقم واحد عند رمي نرد مرة واحدة هو ⅙. وبقية الأحداث لها احتمالات يمكن الاستدلال عليها استنادًا إلى بيانات سابقة. حيث تبلغ احتمالات نجاح ركلة النقطة الواحدة بعد إحراز الهدف في مباريات كرة القدم الاحترافية ٩٤. وهو ما يعني أن الراكلين يحرزون ما متوسطه ٩٤ من كل محاولة لإحراز تلك النقط الوحيدة. (بالطبع قد يختلف هذا الرقم قليلًا لدى راكلين مختلفين في ظل ظروف مناخية مختلفة وظروف أخرى غيرها، لكنها لن تختلف بشكل جذري). يتلخص الأمر بكل بساطة في حقيقة أن امتلاك هذه المعلومات واعتبارها يمكننا من أن نجعل عملية اتخاذ القرار تتسم بالوضوح، وأن يوضح المخاطر التي تحفُها. فمثلًا قد نشر مجلس سلامة النقل الأسترالي تقريرًا يقدر عدد المخاطر المميتة التي تحف كل وسيلة من وسائل النقل. ورغم شيوع الخوف من الطيران فإن المخاطر المرتبطة بالنقل الجوي محدودة جدًا. فلم تشهد أستراليا أي حادثة مميتة في هذا المجال منذ ستينيات القرن العشرين؛ ولهذا فإن معدل الوفيات لكل ١٠٠ مليون كيلومترًا يبلغ صفرًا. أما معدل وفيات سائقي السيارات لكل ١٠٠ مليون كيلومترًا يقطعونه فيبلغ ٥. أما الرقم المذهل في هذا الصدد فهو لراكبي الدراجات النارية، وذلك إذا ما كنت قد ألهمت فقررت أن تتبرع بأعضائك بعد الوفاة. فمعدل الضحايا لدى راكبي الدراجات النارية يبلغ خمسة وثلاثين مِثلًا لسائقي السيارات[٢].

في سبتمبر من عام ٢٠١١ كان قمر صناعي تابع لوكالة ناسا يبلغ وزنه ٦,٥ أطنان يهبط بسرعة كبيرة تجاه الأرض، وكان من المتوقع أن يتفتت هذا القمر الصناعي إلى شظايا ما إن يرتطم بالغلاف الجوي الأرضي. كم كانت احتمالات أن يضربني حطام ذلك القمر؟ أكان يجدر بي أن أُبقي الأطفال في المنزل ولا أذهب بهم إلى المدرسة؟ قدَّر العلماء احتمالات أن يضرب حطام ذلك القمر الساقط أي فرد بـ ١ لـ ٢١ تريليون. لكن احتمالات أن يضرب هذا الحطام أي شخص على وجه الأرض هو ١ لكل ٣٢٠٠ *. ما حدث بعدها أن القمر الصناعي قد تحطم لدى دخوله الغلاف الجوي، لكن العلماء لم يتوصلوا عن يقين إلى أماكن تساقط الحطام[٤]؛ إذ لم تسجل أي إصابات سببها ضربة من هذا الحطام. لا تنبئنا الاحتمالات بما سيحدث يقينًا، وإنما تخبرنا بما يرجح حدوثه وما لا يرجح حدوثه. والراشدون من الناس قادرون على استعمال كل صنوف الأرقام تلك في مجال الأعمال وفي كل مناحي الحياة كذلك. فإنك عندما تسمع أن هناك قمرًا صناعيًّا يهبط بتهور قاصدًا الأرض، مثلًا، فلا يجدر بك أن تندفع على دراجتك البخارية كي تحذر ذويك.

عندما يتعلق الأمر بالمخاطر لا تتماشى الأرقام دائمًا مع ما تطلب منا أن نخافه. في كتاب Freakonomics لكاتبيه ستيف ليفيت وستيفن دوبنر، نجد اكتشافًا مذهلًا مفاده أن برك السباحة الموجودة في ساحات المنازل الخلفية أكثر خطورة من المسدسات التي يخبئها الناس في خزاناتهم[٥] قام ليفيت ودوبنر بحسابات توصلا من خلالها إلى أن أي طفل دون العاشرة من عمره معرض لأن يلقى حتفه في بركة سباحة أكثر من أن يقتل جراء حادث إطلاق نار عن طريق الخطأ بنسبة ١ إلى ١٠٠**. وفي دراسة مثيرة للاهتمام أجراها جاريك بلالوك وفريندا كاديالي ودانييل سايمون، الباحثون في جامعة كورنيل، اتضح أن آلاف الأمريكيين الذين توفوا عقب هجمات الحادي عشر من سبتمبر قد توفوا لأنهم يخشون الطيران[٦] فلن نتيقن أبدًا إذا ما كان في الطيران مخاطرة حقيقية نابعة من إرهاب، لكننا موقنون بأن قيادة السيارات هي أمر مفعم بالمخاطر. ففي عام ٢٠٠١، وبعد أن عزف الأمريكيون عن الطيران وفضلوا السفر بسياراتهم بعد الحادي عشر من سبتمبر، قُدِّرت زيادة في عدد حوادث السيارات خلال أشهر أكتوبر ونوفمبر وديسمبر بعدد ٣٤٤ حادثة (مع الأخذ في الاعتبار متوسط عدد الوفيات والعوامل الأخرى التي تساهم في

* أوضحت ناسا كذلك أنه حتى هذا الحطامُ الساقط من السماء يعتبر ملكية للحكومة. والواضح أنه من غير القانوني أن يحتفظ أحدهم بتذكار من هذا القمر الصناعي، حتى لو سقط ذلك التذكار في ساحة منزلك الخلفية.

** كانت حسابات ليفيت ودوبنر كما يلي: في كل عام يغرق نحو ٥٥٠ طفلًا دون العاشرة، بينما يُقتل ١٧٥ جرَّاء حوادث إطلاق نار عن طريق الخطأ. وكانت المعدلات التي قارنوا بينها هي حالة غرق واحدة لكل ١١٠٠٠ بِركة سباحة منزلية، في مقابل حالة وفاة جرَّاء إطلاق نار خطأ لكل «مليون أو ما يزيد» من المسدسات. لكنني أتوقع أن تختلف الأرقام كثيرًا بالنسبة للمراهقين؛ وذلك لأنهم أكثر مهارة في السباحة، وكذلك لأنهم أكثر قابلية لإحداث مأساة إذا ما وقع في أيديهم سلاحٌ محشوٌ. لكنني لم أتفحص البيانات المتعلقة بهذه النقطة.

حوادث الطرق كحالة الطقس مثلًا). لكن هذا التأثير أخذ يتبدد بمر الزمن، ربما بسبب انحسار خطر الإرهاب، لكن من قاموا بتلك الدراسة وضعوا تقديرًا للحوادث التي سبَّبتها هجمات سبتمبر بألفَي حادثة.

كما أن الاحتمالات قادرة في بعض الأحيان على أن تخبرنا، بعد وقوع الواقعة، بما يرجح أنه حدث وما يرجح أنه لم يحدث – كما في تحليل الحمض النووي. فعندما يكتشف الفنيون في مسلسل سي إس آي ميامي أثر لعاب على بقايا تفاحة قرب ضحية جريمة قتل، فهذا اللعاب لا يحمل اسم القاتل حتى عندما يضع هؤلاء الفنيون، بهيُو الطلعة، هذا اللعاب تحت مجهر قوي. وإنما ما يحدث أن اللعاب (أو الشعرة أو قطعة الجلد أو شظية العظام) يحتوي على مقتطع من مقتطعات الحمض النووي. وكل مقتطع يحمل بدوره مناطق أو نقاط تتباين من فرد لآخر (عدا لدى التوائم المتطابقة الذين يشتركون في نفس الحمض النووي). وعندما يكتب الفاحص الطبي تقريره مؤكدًا «تطابق» عينة الحمض النووي، يعتبر هذا جزءًا مما يتعين على الادعاء إثباته. نعم لا بد أن تتطابق النقاط التي فحصت في عينة الحمض النووي التي وجدت في مسرح الجريمة مع نقاط عينة الحمض النووي التي أخذت من المشتبه به، لكن يتعين على المدعين أن يثبتوا أن ذلك التطابق بين عينتَي الحمض النووي لم يكن محض صدفة.

يشترك البشر في تشابهات في حمضهم النووي تمامًا كما يتشابهون في أمور أخرى مثل مقاسات الأحذية والطول ولون العينين. (أكثر من ٩٩ بالمائة من الحمض النووي يتطابق لدى جميع البشر). فلو لم يتمكن الباحثون إلا من بحث عينة صغيرة من الحمض النووي لا يفحصون فيها إلا القليل من النقاط، فمن الممكن حينها أن يشترك الآلاف بل والملايين من الأفراد في تلك الشظية الجينية. وهكذا فإنه كلما زادت النقاط المفحوصة وزادت التباينات الجينية الطبيعية في كل نقطة من تلك النقاط، كان التطابق أكثر تأكيدًا؛ أو بعبارة أخرى تنخفض احتمالات تطابق تلك العينة مع أكثر من شخص[٧]

كي أسهِّل هذا الأمر، لك تخيل «أن رقم الحمض النووي خاصتك» يتألف من رقم هاتفك مضافًا إليه رقم الضمان الاجتماعي الخاص بك. هذا التتابع المكون من تسعة عشر رقمًا يعرفك بشكل متفرد. الآن لنفترض أن كل رقم منها هو «خانة» ذات عشرة احتمالات وهي: ٠، ١، ٢، ٣، وهلم جرًّا. والآن افترض أن محققي مسرح الجريمة قد وجدوا بقايا «رقم حمض نووي» بهذا الشكل: --٩٥٤---٤-٠-٩٨١٧---. وهو الرقم الذي تصادف أنه يتطابق بالضبط مع «رقم الحمض النووي الذي يخصك». هل أنت مذنب؟

لا بد لك أن تراعي ثلاثة أمور؛ أولها أن ما دون التطابق التام للجينوم كاملًا يترك مجالًا لعدم اليقين، وثانيها أنه كلما زاد عدد «النقاط» التي يمكن فحصها انخفضت نسبة عدم اليقين بالتطابق، وثالثها أن السياق مهم؛ فهذا التطابق سيكون دالًّا جدًّا إذا ما كنت

قد شوهدت تفر مسرعًا من مسرح الجريمة حاملًا بطاقة ائتمان الضحية في جيبك.

عندما يتاح للباحثين الوقت والموارد، تتضمن العملية النموذجية فحص ثلاث عشرة نقطة مختلفة. فاحتمالات أن يشترك شخصان في ثلاث عشرة نقطة متطابقة في الحمض النووي ضعيفةٌ للغاية. عندما استُخدم الحمض النووي للتعرف على الرفات الذي وُجد في مركز التجارة العالمي عقب أحداث الحادي عشر من سبتمبر، وُجد أن العينات التي عثر عليها في موقعٍ ما متطابقة مع العينات التي قدَّمها ذوو الضحايا. كانت الاحتمالات المطلوبة لحدوث تحديد إيجابي هي واحدًا من كل مليار أو أقل، بمعنى أن احتمالات كون الرفات المكتشَف يخص شخصًا غير الضحية المحددة لا بد أن يكون واحدًا من كل مليار أو أقل. في وقت متقدم من هذا البحث أُرخي هذا المعيار بعد أن وجدت جثث أقل لضحايا مجهولين يمكن أن لا نتيقن من تطابقها.

لكن عندما تكون الموارد محدودة، أو تكون عينة الحمض النووي المتاحة صغيرة جدًّا أو ملوثة جدًّا لدرجة لا تسمح بفحص نقاطها الثلاث عشرة، يحتدم الجدل ويغدو الأمر مثيرًا للاهتمام. في عام ٢٠٠٨ نشرت صحيفة لوس آنجيليس تايمز سلسلة تحقيقات صحفية تبحث فيها في استخدام الحمض النووي كدليل جنائي[٨]. وقد تناولت التايمز على وجه الخصوص مسألة إذا ما كانت الاحتمالات التي تستخدمها قوات إنفاذ القانون تستخف بقابلية التطابق عن طريق الصدفة. (إذ بما أنه لا أحد يعرف رسم الحمض النووي لجميع البشر، فإن الاحتمالات التي يقدمها الـ إف بي آي وغيره من جهات إنفاذ القانون للمحكمة هي محض تقديرات). لكن هذه النظرية تعرضت لنكبة علمية عندما كان أحد المحللين بمعمل جنائي في أريزونا يفحص اختبارات عبر قاعدة بيانات الحمض النووي بالولاية، فاكتشف اثنين من المذنبين لا تربطهما صلة قرابة، ورغم هذا يتطابق حمضهما النووي في تسع نقاط، في حين كان الـ إف بي آي يزعم أن احتمالات حدوث تطابق في تسع نقاط تبلغ ١ من كل ١١٣ مليار. ثم اكتشفت أبحاث لاحقة جرت على قواعد بيانات حمض نووي أخرى أكثر من ألف حالة تطابق بين أزواج من البشر في تسع نقاط أو أكثر. سأترك هذه القضية لجهات إنفاذ القانون ومحامي الدفاع ليتعاملوا معها. أما الدرس المستفاد الآن فمفاده أن العلم الكامن وراء تحليل الحمض النووي لا يصح إلا بصحة الاحتمالات التي تدعمه.

من المهم دائمًا أن نعرف أرجحية وقوع الأحداث. ما احتمالات انقطاع التيار الكهربي مع عطل المولد الاحتياطي؟ إن وقوع حدثين مستقلين هو نتاج لاحتمالات كل منهما متعاقبين. بعبارة أخرى نقول إن احتمالات وقوع الحدث أ والحدث ب هي احتمالات وقوع الحدث أ مضروبة في احتمالات وقوع الحدث ب. لمزيد من التوضيح نسوق المثال التالي. إذا كانت احتمالات رمي العملة السليمة فتسقط كاشفة وجه الرأس هي ½، فإن

احتمالات وقوع هذا الحدث مرتين متتاليتين هي ½ × ½ أو ¼. واحتمالات وقوعه ٣ مرات متتالية هي ⅛، واحتمالات وقوعه أربع مرات متتالية هي ١/١٦، وهلم جرًّا. (وبالطبع تدرك أن احتمالات وقوع العملة كاشفة وجه الذيل أربع مرات متتالية هي أيضًا ١/١٦). كما أن ذلك يفسر سبب إصرار مسئولي نظام الكمبيوتر في مدرستك أو في مكتبك على أن يجعلوك تجعل كلمة السر الخاصة بحسابك على النظام «أكثر جودة». فإنك إذا اخترت كلمة سر مكونة من ستة رموز كلها أرقام، فإننا قادرون على أن نحسب عدد كلمات السر الممكنة على غرار: ١٠×١٠×١٠×١٠×١٠×١٠، وهو ما يساوي ١٠٦ أو ١٠٠٠٠٠٠، وهو ما قد يبدو كاحتمالات كثيرة، لكن الكمبيوتر، في الواقع، قادر على أن يمخر عباب تلك التوليفات الـ ١٠٠٠٠٠٠ في غضون جزء من الثانية.

لذا دعنا نفترض أن مسئولي نظام الكمبيوتر ظلوا يزعجونك حتى اضطررت إلى أن تضمِّن أحرفًا في كلمة السر التي تستخدمها. هنا يصير لكل واحد من الرموز الستة ٣٦ توليفة تتكون من ٢٦ حرفًا و١٠ أرقام. وسيزيد عدد كلمات السر الممكنة إلى ٣٦×٣٦×٣٦×٣٦×٣٦×٣٦ أو ٣٦٦ وهو العدد الذي يتجاوز المليارين. أما إذا طالبك المسئولون بأن تجعلها مؤلفة من ثمانية رموز من بينها رموز مثل # و@ و٪ و! كما تفعل جامعة شيكاجو، حينها سيقفز عدد الكلمات الممكنة إلى ٤٦٨ أو ما يتجاوز الـ ٢٠ تريليون.

لكن هناك اختلافًا مهمًّا، فهذه المعادلة لا تصير قابلة للتطبيق إلا إذا كان الحدثان منفصلين، أي أن يكون ناتج أحدهما غير ذي تأثير على ناتج الآخر. فمثلًا ليس لاحتمالات أن ترمي العملة فتسقط كاشفة لوجه الرأس في المرة الأولى أي تأثير على احتمالات سقوطها كاشفة وجه الرأس في المرة الثانية. لكن في المقابل لا تنفصل احتمالات أن تمطر السماء اليوم عما إذا كانت قد أمطرت بالأمس أم لا، فالجهات الهوائية يمكن أن تستمر لأربعة أيام. وبالمثل لا تنفصل احتمالات أن تتصادم سيارتك في حادثة اليوم عن احتمالات أن تتصادم مرة أخرى العام المقبل. فأيًّا ما كان السبب في اصطدامك اليوم، قد يتكرر مرة أخرى العام القادم؛ فقد تكون ميالًا لأن تقود وأنت سكران، أو تنجرف في سباقات مجنونة بسيارتك، أو أن تراسل الناس نصيًّا بينما تقود السيارة، أو أنك مجرد قائد سيء (ولهذا ترتفع أقساط التأمين التي تدفعها لسيارتك بعد تعرُّضك لحادثة، وليس ذلك لمجرد أن شركة التأمين تريد استرداد الأموال التي دفعتها جراء الحادثة، وإنما لأنها امتلكت معلومات جديدة عن احتمالات تعرُّضك لحادثة أخرى مستقبلًا، وهي الاحتمالات التي – بعد أن اصطدمت بباب مرآب منزلك – قد ارتفعت).

لنفترض أنك مهتم باحتمالات وقوع أحد الحدثين، أي الناتج أ أو الناتج ب (مرة أخرى مع افتراض انفصالهما أحدهما عن الآخر). إذا كانت الأحداث تتسم بحصرية متبادلة، كما في إحراز رقم ١ أو ٢ أو ٣ برمية نرد واحدة، فإن احتمالات وقوع الحدث

أ أو ب تتألف من مجموع احتمالاتهما المستقلة: احتمالات الحدث أ مضافة إلى احتمالات الحدث ب. فمثلًا تتمثل أرجحية إلقاء النرد على رقم ١ أو ٢ أو ٣ من رمية واحدة في مجموع احتمالاتهم المستقلة: $\frac{1}{6}+\frac{1}{6}+\frac{1}{6}=\frac{3}{6}=\frac{1}{2}$. وهذا ما يجعل الأمر يبدو منطقيًّا بديهيًّا. عندما يرمى النرد ينتج عن ذلك نتيجة من ست نتائج ممكنة. وتمثل الأرقامُ ١ و٢ و٣ مجتمِعةً نصفَ هذه النتائج الممكنة. ومن ثم فإن فرصتك لأن ترمي النرد على أرقام ١ أو ٢ أو ٣ تبلغ ٥٠ بالمائة. مما يعني أنك إذا ما كنت تلعب لعبة الكرابس في لاس فيجاس، فإن فرصتك في أن ترمي نردين معًا فتصيب رقمي ٧ أو ١١ تمثل عدد التوليفات الرقمية التي مجموعها ٧ أو ١١ في رمية واحدة مقسومًا على عدد التوليفات الرقمية التي يمكن أن تنتج عن رمي نردين أو $\frac{8}{36}$ *.

أما إذا لم تكن الأحداث تشترك في حصريتها، كما في سحب ورقة الخمسة أو ورقة القلب من مجموعة أوراق اللعب، فإن احتمالات وقوع الحدث أ أو الحدث ب تتكون من احتمالات وقوع كل منهما منفردًا مطروحًا منهما احتمالات وقوعهما مجتمعين. مرة أخرى يبدو الأمر منطقيًّا بديهيًّا. في مجموعة أوراق اللعب ٥٢ ورقة. تبلغ احتمالات سحب ورقة الخمسة ٥٢/٤. أما احتمالات سحب ورقة القلب فتبلغ ٥٢/١٣. لكننا لا نستطيع أن نجمع احتمالات الاثنين معًا؛ وذلك بكل بساطة لأن هناك من ورقات القلوب ما هي أيضًا ورقة خمسة (والعكس صحيح). وهكذا فاحتمالات سحب ورقة خمسة أو ورقة قلب هي احتمالات سحب ورقة خمسة (٥٢/٤) مجموعًا لها احتمالات سحب ورقة القلب (٥٢/١٣) مطروحًا منهما احتمالات سحب ورقة القلب الخماسية (٥٢/١) أو ٥٢/١٦.

يمكننا أن نفكر في الأمر بشكل مغاير قليلًا، فنرى: كم من أوراق المجموعة تندرج تحت مسمى الخمسة أو مسمى القلب؟ أولًا هناك ثلاث عشرة ورقة قلب في المجموعة، بعد أن حسبنا تلك الأوراق الثلاث عشرة تتبقى لدينا الأوراق التي هي إما قلب وإما خمسة، والتي هي الخمسة البستونية والخمسة الديناربة والخمسة السباتية. أما الخمسة القلب فقد أحصيناها مسبقًا. مرة أخرى نجد لدينا ١٦ سحبة ممكنة من بين ٥٢ ورقة.

كما أن الاحتمالات تمكننا من حساب ما قد يكون أكثر الأدوات نفعًا في مجال اتخاذ

---

* هناك ستة طرق لإصابة الرقم ٧ برمي نردين: (١، ٦) و(٢، ٥) و(٣، ٤) و(٦، ١) و(٥، ٢) و(٤، ٣). في حين لا توجد سوى طريقتين لإصابة الرقم ١١ وهما: (٥، ٦) و(٦، ٥).

لكن هناك ٣٦ نتيجة ممكنة لرمي نردين وهي: (١، ١) و(١، ٢) و(١، ٣) و(١، ٤) و(١، ٥) و(١، ٦) و(٢، ١) و(٢، ٢) و(٢، ٣) و(٢، ٤) و(٢، ٥) و(٢، ٦). و(٣، ١) و(٣، ٢) و(٣، ٣) و(٣، ٤) و(٣، ٥) و(٣، ٦). و(٤، ١) و(٤، ٢) و(٤، ٣) و(٤، ٤) و(٤، ٥). و(٤، ٦) و(٥، ١) و(٥، ٢) و(٥، ٣) و(٥، ٤) و(٥، ٥) و(٥، ٦). وأخيرًا (٦، ١) و(٦، ٢) و(٦، ٣) و(٦، ٤) و(٦، ٥) و(٦، ٦).

ومن ثم فإن فرصة إصابة الرقم ٧ أو ١١ هي عدد الطرق الممكنة لإصابة واحد من هذين الرقمين مقسومًا على مجموع الرميات المممكنة لنردين، والذي هو ٣٦/٨. والمصادفة هنا أن شطرًا كبيرًا من الأبحاث التي أجريت على هذه الاحتمالات قام بها مقامرون يحاولون تحديد فرصهم في المكسب.

القرار الإداري وفي التمويل على وجه التحديد، ألا وهي: القيمة المتوقعة. فالقيمة المتوقعة تخطو بالاحتمالات الأساسية خطوة إضافية إلى الأمام. إن القيمة المتوقعة أو الربح المتوقع من حدث ما، كشراء تذكرة يانصيب مثلًا، هي مجموع كل النواتج المختلفة كل منها يوزن من خلال احتمالاته أو الربح المتوقع منه. كالعادة يتضح الأمر بالأمثلة. افترض أنك دُعيت إلى لعبة ترمي فيها نردًا واحدًا. الربح المتوقع من هذه اللعبة هو دولار واحد إذا ما أصبت الرقم ١، ودولاران إذا أصبت الرقم ٢، وثلاثة دولارات إذا أصبت الرقم ٣، وهلم جرًّا. ما القيمة المتوقعة من رمية واحدة للنرد؟ كل نتيجة متوقعة لتلك الرمية لها احتمالات تقدر بـ $1/6$، إذن فالقيمة المتوقعة هي: $1/6$ (دولار) + $1/6$ (دولارين) + $1/6$ (ثلاثة دولارات) + $1/6$ (أربعة دولارات) + $1/6$ (خمسة دولارات) + $1/6$ (ستة دولارات) = $\frac{21}{6}$ أو ٣,٥٠ دولارات.

للوهلة الأولى قد يبدو الرقم المعبر عن القيمة المتوقعة ٣,٥٠ دولارات رقمًا غير ذي نفع نسبيًّا. فإنك لا تستطيع أن تجني ٣,٥٠ دولارات من رمية واحدة (لأن ربحك لا بد أن يكون رقمًا صحيحًا). لكن القيمة المتوقعة تتمتع في الواقع بقوة كبيرة؛ لأنها تنبئك بما إذا كان حدثٌ ما «معقولًا» نظرًا إلى تكلفته والناتج المتوقع منه. افترض أن فرصة أتيحت لك كي تلعب تلك اللعبة التي ذكرناها سابقًا في مقابل ثلاثة دولارات. هل يعتبر لعب هذه اللعبة منطقيًّا؟ نعم، لأن القيمة المتوقعة من الناتج هي (٣,٥٠ دولارات) تزيد عن تكلفة اللعبة وهي (٣ دولارات). لا يضمن هذا لك أن تجني المال من لعبة مفردة، لكنه يساعد على توضيح أي المخاطر تستحق المجازفة وأيها لا تستحق.

يمكننا أن نأخذ مثالًا افتراضيًّا ونطبقه على كرة القدم الاحترافية. أشرنا سابقًا إلى أن فريق كرة القدم بعد إحراز الهدف يجد نفسه أمام اختيارين؛ إما أن يركل ركلة النقطة الوحيدة، وإما أن يلعب اللعبة ذات النقطتين. تتطلب ركلة النقطة الواحدة ركل الكرة بين عارضتي المرمى من عند خط الثلاث ياردات، أما لعبة النقطتين فتتطلب العدو والمرور خلال منطقة النهاية من عند خط الثلاث ياردات، وهو الأمر الأكثر صعوبة بكثيرة. يستطيع الفريق أن يتبنى الخيار الأسهل ويحصل على النقطة الوحيدة، أو يمكنهم أن يلجئوا إلى الخيار الأصعب ويحصلوا على النقطتين. ما العمل؟

قد لا يمارس الإحصائيون كرة القدم ولا يواعدون فتيات التشجيع، لكنهم قادرون على تقديم الإرشادات الإحصائية لمدربي اللعبة[٩] وكما قلنا سابقًا تبلغ احتمالات نجاح الركلة عقب إحراز الهدف ٩٤. وهذا يعني أن القيمة المتوقعة لإحراز النقطة عقب المحاولة هي أيضًا ٩٤. بما أنها تساوي الربح المتوقع (نقطة واحدة) مضروبًا في احتمالات النجاح (٩٤). لا يمكن أن يحرز أي فريق ٩٤ نقطة، لكن هذا الرقم يساعد في تحديد قيمة محاولة اتخاذ هذا الخيار عقب إحراز الهدف في مقابل البديل، والذي هو اللعبة ذات النقطتين.

فالقيمة المتوقعة من «محاولة إحراز النقطتين» أقل بكثير إذ تبلغ ٠,٧٤ نعم، الربح المتوقع أعلى حيث يبلغ (نقطتين)، إلا أن معدل النجاح أقل بكثير إذ يبلغ (٠,٣٧). لكن بالطبع إذا كان الوقت المتبقي من المباراة بعد إحراز الهدف هو ثانية واحدة، فلا مناص من المضي إلى اللعبة ذات النقطتين، لكن إذا كان الفريق يهدف إلى تعظيم إجمالي عدد الأهداف على مر الزمن، فإن السبيل الأمثل لفعل ذلك هو الركلة ذات النقطة الوحيدة.

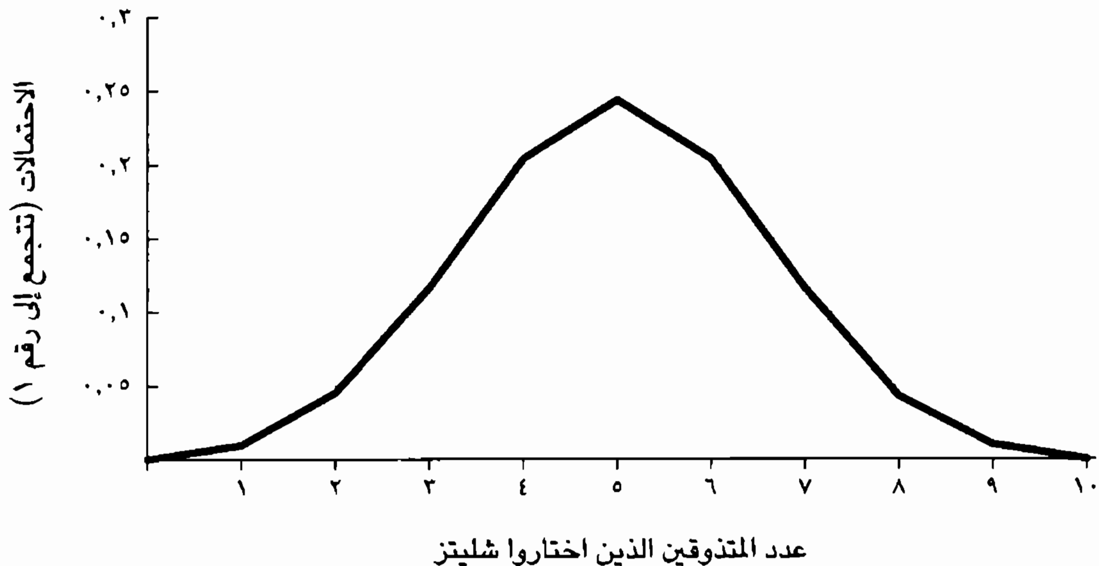
نفس هذا التحليل الأساسي قادر على تبيان السبب الذي لأجله لا ينبغي لك أن تبتاع تذكرة يانصيب. في إيلينوي تطبع احتمالات الربح المتوقع الممكنة من اللعبة على ظهر كل تذكرة. وقد قمت أنا بشراء تذكرة فورية من فئة الـ ١ دولار. (سؤال إلى نفسي: هل ينطبق الخصم الضريبي على هذا السعر؟). على ظهر التذكرة كتبتُ، بخط متناهي الصغر، الفرص الممكنة لربح جوائز مالية مختلفة، أو ربح تذكرة أخرى مجانية، ١ إلى ١٠ (ربح تذكرة مجانية)، ١ إلى ١٥ (ربح دولارين)، ١ إلى ٤٢,٨٦ (ربح أربعة دولارات)، ١ إلى ٧٥ (ربح خمسة دولارات)، وهلم جرًّا حتى احتمالات تبلغ ١ إلى ٤٠٠٠٠ فرصة لربح ١٠٠٠ دولار. قمت بحساب الربح المتوقع من تذكرتي الفورية عن طريق جمع كل الجوائز المالية الممكنة مقيِّمًا إياها من خلال احتمالاتها*. فاتضح أن تذكرتي ذات الدولار الواحد لها ربح متوقع يقدر بنحو ٠,٥٦ دولار، مما يجعل من ينفق دولارًا لأجلها شخصًا بائسًا. ولأنني رجل محظوظ فقد ربحت دولارين.

لكن رغم ربحي لهذين الدولارين كان شرائي لهذه التذكرة عملًا أخرق. ذلك واحد من أهم الدروس المستفادة من الاحتمالات؛ ألا وهو أن القرارات الجيدة – قياسًا لما تنطوي عليه من احتمالات – من الممكن أن تئول مآلًا سيئًا. وأن القرارات السيئة – كإنفاق دولار على شراء تذكرة يانصيب بإيلينوي – من الممكن أن تئول مآلًا جيدًا على المدى القصير على الأقل. لكن الاحتمالات هي التي تنتصر في النهاية. هناك نظرية مهمة تعرف باسم قانون الأرقام الكبيرة، تنص على أنه كلما زاد عدد المحاولات المنفصلة، اقترب متوسط النواتج من

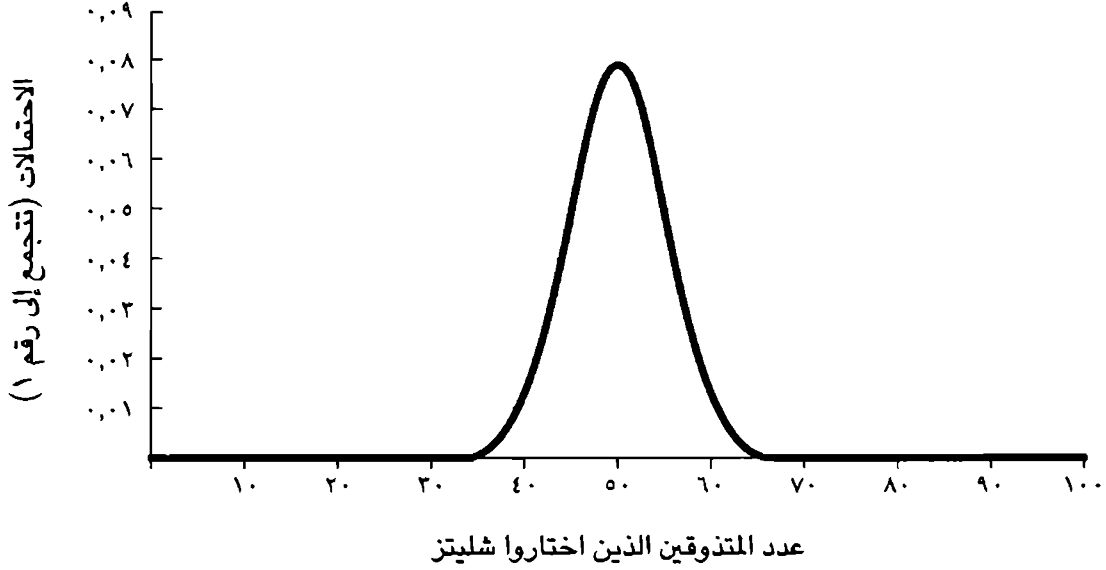
* القيم المتوقعة من تذاكر ياناصيب دوجوت دابلر بإيلينوي ذات الدولار الواحد (مقربة إلى سنتات) هي كما يلي: ١/ ١٥ (دولاران) + ١/ ٤٢,٨٦ (أربعة دولارات) + ١/ ٧٥ (خمسة دولارات) + ١/ ٢٠٠ (عشرة دولارات) + ١/ ٣٠٠ (خمسة وعشرون دولارًا) + ١/ ١٥٨٩,٤٠ (خمسون دولارًا) + ١/ ٨٠٠٠ (مائة دولار) + ١/ ١٦٠٠٠ (مائتا دولار) + ١/ ٤٨٠٠٠ (خمسمائة دولار) + ١/ ٤٠٠٠٠ (ألف دولار) = ٠,١٣ دولار + ٠,٠٩ دولار + ٠,٠٧ دولار + ٠,٠٥ دولار + ٠,٠٨ دولار + ٠,٠٢ دولار + ٠,٠١ دولار + ٠,٠١ دولار + ٠,٠٢ = ٠,٥٢ دولار. لكن هناك أيضًا فرصة تبلغ ١/ ١٠ للحصول على تذكرة مجانية، وهي فرصة ذات ربح متوقع يبلغ (نحو) ٠,٥٢ دولارًا. وبنوع من التقريب نجد إجمالي الربح المتوقع يساوي٠,٥٢ دولار + ١ (٥٢) = ٠,٥٧ دولار. سوف يشير أصوليو الرياضيات (الذين يراسلونني كثيرًا عن طريق البريد الإلكتروني) إلى أن هذا غير صحيح. فللتذكرة المجانية فرصة لربح تذكرة مجانية أخرى، وهو ما يتيح قيمة محتملة وإمكانية لربح تذكرة مجانية ثالثة (وهو ما قد يستمر إلى ما لا نهاية). عندما يعالج هذا الأمر بطريقة أكثر تعقيدًا من الناحية الرياضية ترتفع القيمة المتوقعة للتذكرة إلى نحو ٠,٥٨ دولار، وهي القيمة التي تظل سيئة أيضًا.

القيمة المتوقعة. نعم لقد ربحت دولارين اليوم وقد أربح دولارين غدًا، لكنني إذا ابتعت آلاف التذاكر من فئة الدولار الواحد التي لكل واحدة منها ربح متوقع يقدر بـ ٠,٥٦ دولار، هنا تقضي الرياضيات عن شبه يقين بأنني سوف أخسر مالًا. فإذا ما أنفقت مليون دولار على هذه التذاكر لن يعود لي منها إلا مبلغ يقارب ٥٦٠٠٠٠ دولار.

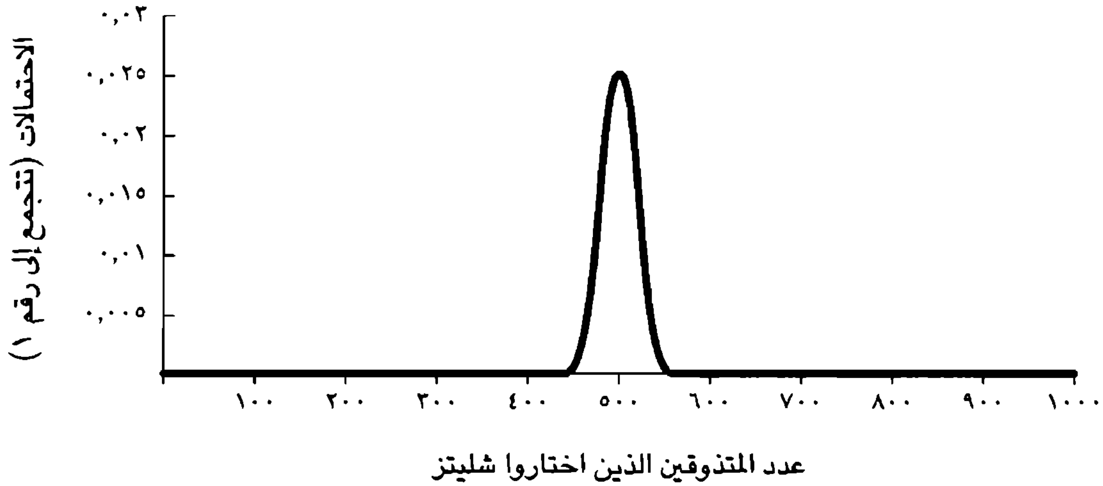
يوضح لنا قانون الأرقام الكبيرة لماذا تجني صالات القمار أموالها على المدى الطويل. فالاحتمالات التي ترتبط بألعاب القمار كلها تصب في صالح الصالات (هذا على افتراض نجاح صالات القمار في منع لاعبي البلاك جاك من إحصاء أوراق اللعب). فإذا ما وضع عدد كافٍ من الرهانات على مدًى زمني كافٍ فسوف تربح الصالة، ولا شك، أكثر مما تخسر. يرينا قانون الأرقام الكبيرة لماذا أحسنت شليتز صنعًا بإجرائها ١٠٠ اختبار تذوق لمعصوبي الأعين خلال استراحة ما بين الشوطين في نهائي البطولة الوطنية لكرة القدم ولم تكتفِ بعشرة اختبارات فقط. فلنتفحص «دالة الكثافة الاحتمالية» لاختبارات من عينة اختبار شليتز بعشر محاولات ومائة محاولة وألف محاولة. (رغم أن الأمر يبدو جميلًا كون دالة الكثافة الاحتمالية لا تفعل سوى أنها تنثر النواتج المتجانسة على طول المحور x والاحتمالات المتوقعة لكل ناتج على طول المحور y، إلا أن الاحتمالات المقيَّمة – كل ناتج مضروب في تواتره المتوقع – سوف تتجمع كلها لتصير رقم ١). مرة أخرى أنا أفترض أن اختبار التذوق كرمي العملة في الهواء، ولكل متذوق احتمالات قدرها ٠,٥ لاختيار شليتز. وكما ترى في الشكل التالي يجمع الناتج المتوقع نحو ٥٠ بالمائة من المتذوقين الذين اختاروا شليتز، بينما يزداد عدد المتذوقين. في الوقت ذاته تنخفض احتمالات الحصول على ناتج ينحرف بشدة عن الـ ٥٠ بالمائة مع ازدياد المحاولات.



١٠٠ محاولة



١٠٠٠ محاولة



قلت سابقًا إن مديري شليتز سيُسَرُّون إذا اختار ٤٠ بالمائة أو أكثر من شاربي ميشيلوب جعة شليتز وهم معصوبو الأعين. الأرقام التالية تظهر احتمالات الوصول إلى هذا الناتج بزيادة عدد المتذوقين:

١٠ متذوقين معصوبو الأعين: ٠,٨٣

١٠٠ متذوق معصوبو الأعين: ٠,٩٨

١٠٠٠ متذوق معصوبو الأعين: ٠,٩٩٩٩٩٩٩٩٩٩

١٠٠٠٠٠٠ متذوق معصوبو الأعين: ١

هنا تتضح المنطقية من وراء عنوان هذا الفصل الفرعي: «لا تشترِ شهادة الضمان

الممتد لطابعتك ذات التسعة والتسعين دولارًا»، أو قد لا تتضح. دعني أوضح الأمر. إن مجال التأمين بأسره قائم على الاحتمالات. (وما شهادة الضمان إلا شكل من أشكال التأمين). فإنك عندما تؤمِّن على شيء ما تتعاقد على أن تتلقى نوعًا من أنواع الربح في حالة محددة من الطوارئ. فمثلًا سيقوم التأمين بإبدالك سيارة بسياراتك إذا ما حدث وأن سُرقت أو سحقتها شجرة ساقطة. لكن في مقابل هذا الضمان توافق أنت على أن تدفع قدرًا ثابتًا من المال طوال الفترة التي تؤمِّن خلالها على سيارتك. الفكرة الأساسية هنا هي أنك قد نقلت خطر سرقة سيارتك، أو انسحاقها أو حتى تلفها نتيجة قيادتك السيئة، إلى شركة التأمين في مقابل دفعات نقدية ثابتة.

لماذا تقبل هذه الشركات تلك المجازفة؟ لأنها ستجني مكاسب هائلة على المدى الطويل إذا ما قامت بتسعير أقساط التأمين بشكل صحيح. بالطبع سوف تتعرض بعض السيارات المؤمَّن عليها لدى شركة آلستايت للسرقة، وسيتلف بعضها الآخر عندما يصدمها سائقوها بصنابير الحريق في الشوارع، كما حدث مع الفتاة التي كنت أواعدها في المدرسة الثانوية (ولقد اضطرت أيضًا لأن تُغيِّر صنبور الحريق، وهو أمر باهظ التكلفة أكثر مما قد تظن). إلا أن غالبية السيارات المؤمن عليها لدى آلستايت أو غيرها من شركات التأمين لن يصيبها ضرر. وما تحتاجه شركات التأمين كي تجني أرباحًا هو أن تحصل من دفعات التأمين ما يفوق ما ستدفعه لإصلاح الأضرار. وفي سبيل ذلك لا بد أن تدرك تلك الشركة جيدًا ما يطلق عليه في رطانة صناعة التأمين «الخسارة المتوقعة» بأي نمط من أنماطه. وهو المفهوم الذي يماثل تمامًا مفهوم «القيمة المتوقعة» مع حيلة تأمينية. فإذا ما أمَّنت على سيارتك بمبلغ ٤٠٠٠٠ دولار، وكانت احتمالات سرقتها في أي سنة من سنوات مدة التأمين تقدر بـ ١ إلى ١٠٠٠، فإن الخسارة السنوية المتوقعة لسيارتك تقدر بأربعين دولارًا. إذن فالقسط السنوي المخصص للتأمين ضد السرقة من إجمالي التغطية التأمينية لا بد أن يتجاوز الدولارات الأربعين. عند هذه النقطة تصير شركة التأمين كصالة القمار أو يانصيب إيلينوي. نعم هناك أموال تدفعها، لكن على المدى الطويل سيفوق الداخلُ المدفوعَ.

وبصفتك مستهلكًا، عليك أن تدرك أن التأمين لن يوفر مالك على المدى الطويل. وإنما ما سيفعله هو أنه سيمنع عنك بعض الخسائر الباهظة التي لا تستطيع تحملها فورًا، كاستبدال سيارة بسيارتك البالغ ثمنها ٤٠٠٠٠ دولار والتي سرقت، أو منزل بمنزلك البالغ ثمنه ٣٥٠٠٠٠ دولار الذي احترق. إن شراء وثيقة التأمين هو «رهان خاسر» من وجهة النظر الإحصائية؛ حيث إنك سوف تدفع لشركة التأمين في المتوسط ما يفوق ما ستأخذ منها. لكنه مع ذلك يمكن أن يكون وسيلة معقولة تحميك مما يمكن أن يدمر حياتك. والمفارقة أن مَن هم في ثراءِ وارن بافيت سيوفرون أموالهم إذا لم يشتروا وثائق تأمين على سياراتهم أو منازلهم أو حتى على صحتهم؛ وذلك لأنهم قادرون على التعامل مع أي ضرر يصيبهم.

وهو ما ينقلنا أخيرًا إلى طابعتك التي ابتعتها بتسعة وتسعين دولارًا. لنفترض أنك انتقيت طابعة ليزرية جديدة ممتازة من شركة بيست باي أو غيرها من شركات البيع بالتجزئة*. وعندما ذهبت لتدفع ثمنها عرض عليك البائع مجموعة من خيارات الضمان الممتد. ففي مقابل ٢٥ أو ٥٠ دولارًا إضافية سوف تقوم شركة بيست باي بإصلاح الطابعة أو تغييرها حال تلفها خلال العام أو العامين التاليين. ولأنك تفهم قواعد الاحتمالات والتأمين والاقتصاد الأساسي، فلا بد أنك قد أدركت فورًا كل ما يلي: (١) بيست باي شركة ساعية إلى الربح وإلى تعظيم الربح. (٢) إن البائع متلهف لأن يجعلك تبتاع شهادة الضمان الممتد تلك. (٣) النقطتان ١ و٢ تدللان لك على أن تكلفة شهادة الضمان أعلى من التكلفة المتوقعة لإصلاح الطابعة في بيست باي. ولولا هذا لما حاولت الشركة بيعك الضمان بهذا الإلحاح. (٤) إذا ما تلفت طابعتك التي سعرها تسعة وتسعين دولارًا واضطررت لأن تنفق على إصلاحها من مالك الخاص، لما أثر هذا على حياتك في شيء.

سوف تكلفك شهادة الضمان، في المتوسط، ما يفوق تكلفة إصلاحها. لكن الفكرة الأعم هنا – وواحد من الدروس الأساسية في مجال التمويل الشخصي – هو أنك تحتاج دومًا لأن تؤمن على نفسك ضد أي طوارئ متعددة لا تستطيع دفع تكاليفها في لحظتها. لكن تجنب التأمين ضد أي شيء آخر.

والقيمة المتوقعة ذات فائدة أيضًا في فك تشابك القرارات المعقدة التي ترتبط بالكثير من الطوارئ التي تقع في مراحل زمنية مختلفة. لنفترض أن صديقًا لك طلب منك أن تستثمر مبلغ مليون دولار في مبادرة بحثية تختبر علاجًا جديدًا لصلع الرجال. سوف تسأله، على الأرجح، عن القابلية المتوقعة لنجاح العلاج لكنك حينها ستتلقى إجابة معقدة. إنه مشروع بحثي مما يعني أن هناك فرصة تبلغ احتمالاتها ٣٠ بالمائة لأن ينجح الفريق في اكتشاف علاج ناجع. وإذا لم ينجح الفريق فسوف يعود لك مبلغ ٢٥٠٠٠٠ دولار، وهو المبلغ الذي كان مدخرًا لتقديم العقَّار للأسواق (عن طريق الاختبار والتسويق وما إلى ذلك). وحتى لو نجح الفريق، فهناك فرصة تبلغ ٦٠ بالمائة فقط لأن تجيز هيئة الغذاء والدواء الأمريكية هذا العقَّار المعجزة وتقضي بأنه آمن للاستخدام البشري. وحتى حينها، حين يثبت أن العقَّار آمن وفعال، فهناك فرصة تبلغ ١٠ بالمائة أن يأتي منافس بعقَّار آخر في نفس الوقت يفوق عقَّارك جودة، جاعلًا كل توقعات أرباحك تذهب أدراج الرياح. أما إذا ما سار كل شيء على ما يرام – العقَّار آمن وفعال ولا ينافسه منافس – فسيربح استثمارك ٢٥ مليون دولار على أفضل تقدير.

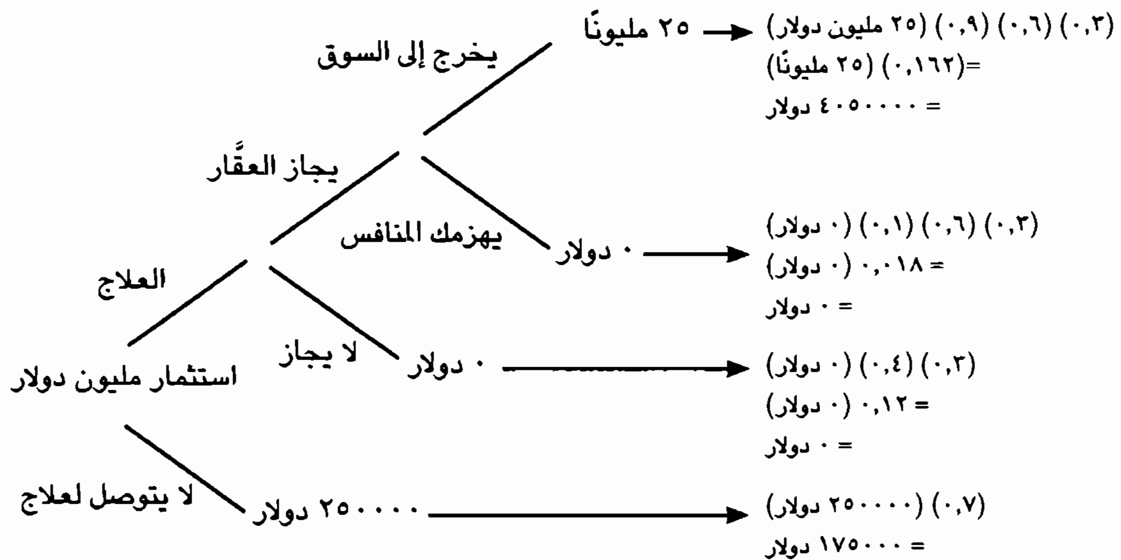
أيجدر بك أن تُقْدم على هذا الاستثمار؟

---

* في موقع سابق من هذا الكتاب أوردتُ مثالًا عن عاملين سكارى يُنتجون طابعات معطوبة. عليك أن تنسى هذا المثال هنا وتتخيل أن تلك الشركة قد أصلحت تلك العيوب.

قد تبدو تلك معلومات مشوشة. فالربح المحتمل كبير – خمسة وعشرون مِثْلًا لما استثمرته – لكن هناك الكثير من المخاطر المحتملة. يمكن لتحليل شجري للقرارات أن يساعد في تنظيم كل هذه المعلومات، وكذلك – إذا ما صحت الاحتمالات المرتبطة بكل ناتج – يمنحك تقييمًا قائمًا على الاحتمالات لما ينبغي عليك فعله. فالتحليل الشجري للقرارات ذلك يرسم لك خريطة لكل ما هو مثير للشك، وكذلك لكل الاحتمالات المرتبطة بالنواتج الممكنة. والجزء الأخير من التحليل الشجري يرينا جميع الأرباح المتوقعة واحتمالات كل واحد منها. وإذا ما قيَّمنا كل ربح طبقًا لأرجحيته ومجموع كل الاحتمالات، فسوف نحصل على القيمة المتوقعة لتلك الفرصة الاستثمارية. وكما هي العادة نجد أفضل الفهم في مطالعة الشكل.

قرار الاستثمار



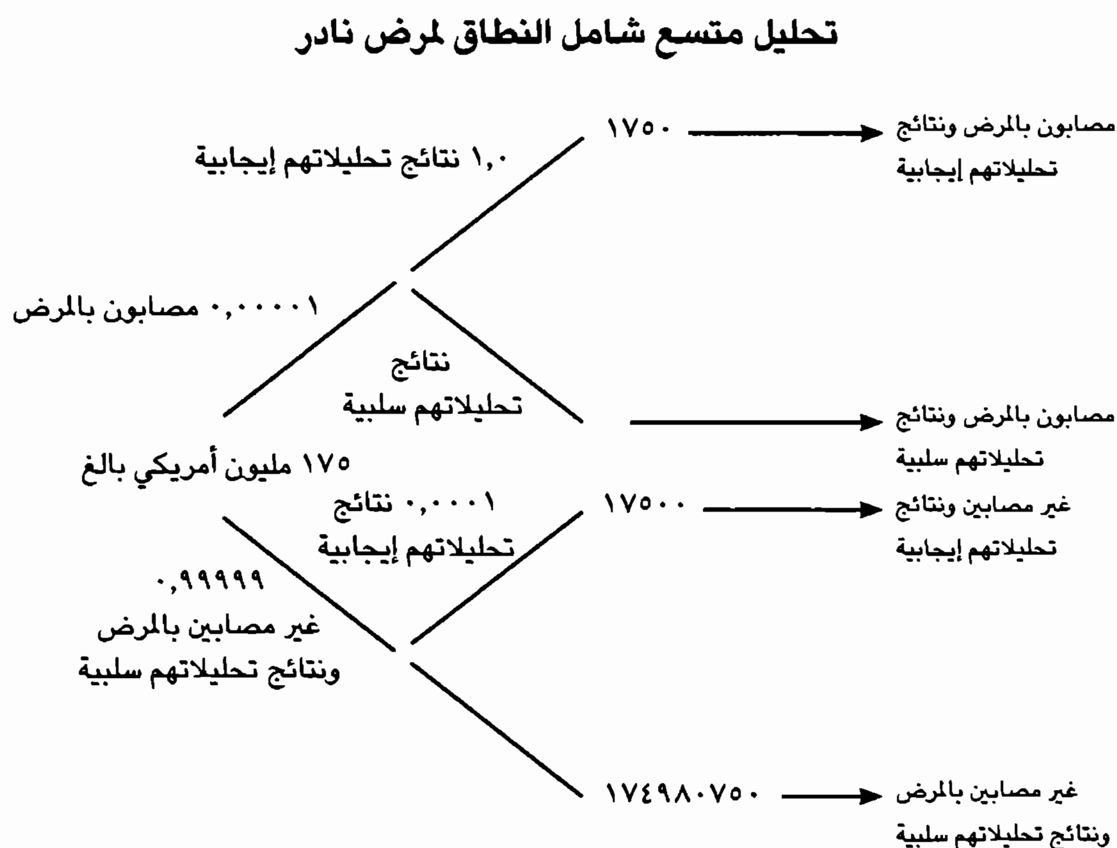
الربح المتوقع = ٤٠٥٠٠٠٠ دولار + ٠ دولار + ٠ دولار + ١٧٥٠٠٠ دولار = ٤٢٢٥٠٠٠ دولار

لهذه الفرصة تحديدًا قيمة متوقعة تسيل اللعاب. فالربح المتوقع ٤,٢٢٥ ملايين دولار. لكن رغم هذا قد لا يكون الاستثمار الأمثل لذلك المال الذي تدخره لدفع الرسوم الدراسية الجامعية لأبنائك. سينبئك التحليل الشجري بأن ربحك المتوقع أعلى بكثير مما يطلب منك ادخاره. لكنك في المقابل ستجد الناتج الأكثر أرجحية، أي الذي يحدث عادة، هو أن الشركة ستفشل في اكتشاف علاج الصلع ذلك ولن يعود لك من مالك سوى ٢٥٠٠٠٠ دولار. تعتمد قابليتك للإقدام على هذا الاستثمار في الغالب على تاريخك مع المجازفة. يفترض قانون الأرقام الكبيرة أن المؤسسات الاستثمارية أو الأفراد الأثرياء مثل وارن بافيت يجدر بهم أن يقدموا على مئات من هذه الفرص ذات النواتج غير الأكيدة والعوائد المتوقعة التي تسيل اللعاب. بعض تلك الاستثمارات سيربح وبعضها سيخفق. لكن هؤلاء المستثمرين سوف يحققون، في المتوسط، الكثير من الأموال تمامًا كما تفعل شركات التأمين وصالات القمار. فإذا كان الربح المتوقع يخدم صالحك، فمن الأفضل أن تقدم على محاولات أكثر.

نفس تلك العملية الأساسية يمكن أن تستخدم لتفسير ظاهرة قد تبدو منافية للمنطق. فقد لا يبدو من قبيل المنطق أن نفحص جميع السكان بحثًا عن مرض نادر رغم فداحته ألا وهو الإيدز/ إتش أي في. لنفترض أننا قادرون على إجراء تحليلٍ بحثًا عن مرض نادر بقدر عالٍ من الدقة. لنفترض في هذا المثال أن هذا المرض يصيب ١ من كل ١٠٠٠٠٠ من البالغين، وأن دقة التحليل تبلغ ٩٩,٩٩٩٩ بالمائة. هذا الاختبار لا يخرج أبدًا تحليلًا سلبيًّا زائفًا (بمعنى أنه لا يخطئ أبدًا في تشخيص شخص مصاب بالمرض)، لكنه في المقابل فيما نسبته ١ من كل ١٠٠٠٠ شخص سليم خضع للاختبار سوف يخرج تحليلًا إيجابيًّا زائفًا، بمعنى أن نتيجة أحد الأشخاص ستأتي موجبة رغم أنه لا يعاني المرض. لكن الناتج المذهل هنا هو أنه رغم دقة التحليل الشديدة فإن غالبية من جاءت نتائج تحليلهم إيجابية لن يصابوا بالمرض. هذا الأمر سيسبب قلقًا عارمًا لدى أولئك الذين جاءت نتائج تحليلهم الزائفة إيجابية، ويمكن أن يؤدي أيضًا إلى استنزاف مواردهم المحدودة والمتعلقة بالرعاية الصحية على العلاجات وفحوصات المتابعة.

وإذا ما اختبرنا جميع البالغين الأمريكيين البالغ عددهم ١٧٥ مليون فرد، فسيبدو التحليل الشجري كما يلي:

تحليل متسع شامل النطاق لمرض نادر



$$\frac{\text{المصابون بالمرض}}{\text{الذين أخبروا بأنهم مصابون بالمرض}} = \frac{١٧٥٠}{١٧٥٠٠ + ١٧٥٠} = \frac{١٧٥٠}{١٩٢٥٠} = ٠,٠٩ = ٪٩$$

المصابون بالمرض هم فقط ١٧٥٠ كلهم نتائجهم إيجابية. لكن ما يزيد عن ١٧٤ مليونًا من البالغين غير مصابين بالمرض. من هذه المجموعة من الأصحاء الذين جاءت نتائجهم إيجابية ٩٩,٩٩٩٩ تلقوا النتائج الصحيحة التي تقضي بأنهم غير مصابين بالمرض. ٠,٠٠٠١ فقط يتلقون نتائج إيجابية زائفة. لكن يظل ٠,٠٠٠١ من الـ ١٧٤ مليونًا رقمًا كبيرًا. بل الواقع أن ١٧٥٠٠، في المتوسط، سوف يتلقون نتائج إيجابية زائفة.

دعنا نتأمل ما يعني ذلك. سوف يتم إبلاغ عدد ١٩٢٥٠ شخصًا بأنهم مصابون بالمرض، لكن ٩ بالمائة منهم فقط هم المرضى بالفعل. وذلك من خلال تحليل يتسم بمعدل منخفض جدًّا للنتائج الإيجابية الزائفة. كي لا ننجرف بعيدًا عن الموضوع نقول إن هذا يمنحك فكرة عن سبب انخفاض ميزانية فحص الأصحاء ضمن الميزانية الكلية للسيطرة على الأمراض. وبالنسبة إلى مرض الإيدز/ إتش آي في ينصح المسئولون بالصحة عادة بتوجيه الموارد المتاحة نحو فحصِ مَن هم أكثر عرضة للإصابة، كالمثليين جنسيًّا من الرجال أو متعاطي المخدرات الوريدية.

• • •

أحيانًا تساعدنا الاحتمالات على أن نتنبه لشركائنا المريبين. في الفصل الأول عرضنا مشكلة الغش المنهج في الاختبارات الموحدة، وعرضنا كذلك لواحدة من المؤسسات التي تقضي عليه، وهي مؤسسة كافيون تيست سيكيوريتي. كما أن لجنة الأوراق المالية والتبادلات (SEC)، والتي هي الوكالة الحكومية المسئولة عن إنفاذ القوانين الفدرالية المتعلقة بتجارة الأوراق المالية، تستخدم منهجية شبيهة للإيقاع بمن يمارسون التجارة الداخلية. (تنطوي التجارة الداخلية على استخدام غير مشروع لمعلومات خاصة كاستخدام معلومات عن مشاريع استحواذ مستقبلية ستقوم بها إحدى المؤسسات في تجارة سندات أو غيرها من الأوراق المالية الخاصة بالشركات التي ستتأثر بالقرار). تستخدم وكالة SEC حواسيب فائقة في فحص مئات الملايين من عمليات تجارة السندات، مفتِّشةً عن نشاط مثير للريبة كشراء عدد كبير من أسهم إحدى الشركات قبل الإعلان عن الاستحواذ عليها مباشرة، أو التخلص من أسهم قبل إعلان الشركة عن خسائرها مباشرة[١٠] كما تحقق الوكالة مع مديري الاستثمار الذين يحظون بعوائد عالية جدًّا على مدًى زمني طويل. (فالنظرية الاقتصادية والبيانات التاريخية كلتاهما تُظهران كم هو من الصعب أن يجني مستثمر واحد عوائد تفوق المتوسط كل عام). بالطبع يحاول أذكياء المستثمرين دائمًا أن يتوقعوا الأنباء الجيدة والسيئة، ومن ثم ينتهجوا استراتيجيات شرعية لمغالبة السوق. فحسن الاستثمار لا يعني بالضرورة أن ممارسه مجرم. كيف للحاسوب أن يجد الفارق إذن؟ لقد هاتفتُ فرع إنفاذ القانون في وكالة SEC عدة مرات كي أسألهم عن تلك الأنماط المحددة التي تشير إلى نشاطات إجرامية، لكنهم لم يردوا عليَّ حتى الآن.

في فيلم تقرير الأقلية Minority Report الذي عرض عام ٢٠٠٢ يلعب توم كروز دور محقق «استباق الجرائم» ضمن مكتب يستخدم التكنولوجيا للتنبؤ بالجرائم قبل أن ترتكب.

الواقع يا رفاق أن ذلك لم يعد خيالًا علميًّا اليوم. ففي عام ٢٠١١ نشرت صحيفة النيويورك تايمز هذا العنوان: «أن تتواجد الشرطة قبل وقوع الجريمة»[١١] يروي المقال كيف أرسل رجال الشرطة إلى مرآب بوسط مدينة سانتا كروز بعد أن تنبأ برنامج كمبيوتر بأرجحية وقوع حوادث سطو بواسطة سيارات موجودة في ذلك الموقع ذلك اليوم. وبناءً على ذلك اعتقلت الشرطة امرأتين كانتا تُنعِمان النظر في نوافذ السيارات. إحداهما كانت تحمل سجلًا حافلًا بمرَّات الاعتقال، أما الأخرى فقد كانت تحمل ممنوعات.

نظام سانتا كروز ذلك صممته مجموعة مؤلفة من عالِمي رياضيات وعالم أنثروبولوجيا وعالم مختص بعلم الجريمة. لقد أنشأت شرطة شيكاجو وحدة تحليلات تنبؤية كاملة، وكان من أسباب هذا كون أنشطة العصابات، التي هي من المصادر الأساسية للعنف الحادث في المدينة، تتخذ أنماطًا معروفة. إذا قرأنا كتاب Data Mining and Predictive Analysis: Intelligence Gathering and Crime Analysis الذي هو دليل إحصائي لجهات إنفاذ القانون، نجده مستهَلًّا بالعبارة الحماسية التالية: «قد صار التنبؤ بالجرائم المستقبلية ممكنًا الآن، فقد صرنا قادرين على أن نحدد الأساليب الإجرامية، وتوقع الأماكن التي ستصير بؤرًا للجريمة في المجتمع، وعلى الدراسة المتأنية للقرارات المتعلقة بنشر القوات وتوفير أفضل حماية للمواطنين بأعلى مستوًى من الكفاءة». (إنني أقرأ هذه الأمور كي لا تضطر أنت لأن تقرأها).

«أعمال الشرطة التنبؤية» هي جزء من حركة أوسع نطاقًا تسمى أعمال التحليل التنبؤية. سوف تظل الجريمة دومًا يتخللها عنصر عدم اليقين، تمامًا كما لا يمكن التيقن من كنه ذلك الشخص الذي سوف يصدم سيارته، أو تلك التي ستتخلف عن سداد أقساط رهن بيتها. تساعدنا الاحتمالات على التعرف على تلك المجازفات. وهذه المعلومات تصقل من إدراكنا للاحتمالات ذات الصلة بتلك المجازفات. فالمؤسسات ذات المستقبل المجهول دائمًا ما تسعى لتقدير المخاطر التي قد تواجهها. والمقرضون دائمًا يطلبون أمورًا على غرار بيان التحقق من الدخل وسجل الوفاء بالديون. لكن أدوات التأمين الفظة تلك قد صارت أقرب للمكافئ التنبؤي للأدوات الحجرية لرجل الكهف. فهذا الحشد الهائل من البيانات الرقمية مع القدرات الحسابية المحدودة قد ولَّدوا أفكارًا مذهلة عن السلوك البشري. يطلق مسؤولو التأمين وصفًا صحيحًا على مجال عملهم مسمِّين إياه «تحويل المخاطر» – ومن ثم فإن عليهم أن يدركوا تلك المخاطر التي ستحول إليهم. وشركات على غرار آل ستايت تتخصص في مجال معرفة أمور تبدو لغيرها ثرثرة تافهة[١٢]

- سائقو السيارات الذين تتراوح أعمارهم ما بين العشرين والرابعة والعشرين هم الأكثر عرضة لحوادث السيارات القاتلة.
- أكثر أنواع السيارات عرضة للسرقة في إيلينوي هي الهوندا سيفيك (وتقابلها في آلاباما الشيفروليه البيك أب العريضة)*.
- المراسلة النصية أثناء القيادة تسبب الحوادث، لكن قوانين الولاية التي تحظرها غير ناجحة في منع السائقين من ممارستها. بل إن تلك القوانين تزيد الطين بلة؛ لأنها تحض السائقين على إخفاء هواتفهم المحمولة، ومن ثم تحويل أنظارهم عن الطريق أثناء كتابة الرسائل.

تعتبر شركات بطاقات الائتمان في طليعة مستخدمي هذا النوع من التحليل؛ وذلك لأنها تمتلك بيانات كثيرة جدًّا عن عاداتنا في الإنفاق، ولأن أنماط أعمالهم تعتمد بشكل كبير على إيجاد العملاء الذين لا يكادون يندرجون تحت مسمى مخاطرين ائتمانيين جيدين. (فالعملاء الذين يُعَدون مخاطرين ائتمانيين ممتازين سوف يخسرون الكثير من الأموال على الأرجح لأنهم يسددون فواتيرهم كاملة كل شهر، أما العملاء الذين يمتلكون أرصدة ضخمة ذات فوائد كبيرة هم من يأتون بعوائد كبيرة إذا لم يتخلفوا عن تسديد أقساطهم). واحدة من أهم الدراسات التي بحثت في موضوع أي الأشخاص أقرب لأن يسدد فواتيره وأيهم أقرب لأن يتملص منها، أجراها جيه بي مارتن، وهو «مدير عاشق للرياضيات» في شركة كاناديان تاير، وهي شركة كبيرة للبيع بالتجزئة تتاجر في منتجات السيارات وغيرها من البضائع[١٢]. عندما قام مارتن بتحليل البيانات – كل عمليات البيع التي أجريت باستخدام بطاقات ائتمان كاناديان تاير في العام السابق – اكتشف أن ما اشتراه الزبائن يعتبر مؤشرًا دقيقًا للسلوك الذي سيسلكونه حيال تسديدهم لأقساطهم فيما بعد عندما يضاف إلى الطرق التقليدية مثل مقدار الدخل وتاريخ الديون.

في مقال نشر بمجلة النيويورك تايمز ماجازين عنوانه: «ماذا تعرف شركة بطاقة الائتمان عنك؟» وردت بعض أهم النتائج التي توصل مارتن إليها على غرار: «أولئك الذين يشترون زيت محرك رخيصًا عاديًّا هم على الأرجح سوف يتخلفون عن تسديد أقساط ديونهم، على خلاف من يشترون زيتًا من نوع فاخر غالي الثمن. ومن يبتاعون أجهزة رصد دخان الحريق أو تلك الوسائد اللبادية التي توضع أسفل أرجل المقاعد كي تحول بينها

* لأنني نصحتك بأن تدقق في الإحصائيات الوصفية، فإنني أجد نفسي مضطرًّا لأن أشير إلى أن أكثر السيارات التي تشيع سرقتها ليست بالضرورة أكثر السيارات التي تحتمل سرقتها. فقد أبلغ عن سرقة الكثير من سيارات الهوندا سيفيك؛ وذلك لأن هناك الكثير منها في الشوارع، أما احتمالات أن تسرق أي سيارة هوندا سيفيك مفردة (وهو ما تهتم به شركات التأمين) منخفضة جدًّا. لكن في المقابل حتى لو سرق ما نسبته ٩٩ بالمائة من سيارات الفيراري لن تصير الفيراري من «أكثر السيارات التي تشيع سرقتها»؛ وذلك لأنه لا يوجد منها الكثير كي يُسرق.

وبين خدش الأرضيات لا يتخلفون عن أي قسط من أقساطهم. لكن من يبتاعون حليات سيارة عى شكل جمجمة مصنوعة من الكروم أو عادم مزود بخاصية دفع كبيرة، فلا بد أنه في النهاية سوف يتخلف عن سداد إحدى فواتيره».

• • •

تمنحنا الاحتمالات أدوات للتعامل مع مجهولات الحياة. لا يجدر بك أن تبتاع تذاكر اليانصيب. ويجدر بك أن تستثمر في سوق الأوراق المالية إذا كان أفقك الاستثماري طويل المدى (لأن الأوراق المالية تأتي بعوائد على المدى الطويل). عليك أن تؤمِّن على أشياء معينة ولا تؤمن على غيرها. بل إن الاحتمالات قادرة على أن تساعدك في الربح في بعض برامج المسابقات (كما سترى في الفصل التالي).

بعد أن قلنا ما سبق (أو كتبناه) لا بد أن نقول إن الاحتمالات ليست بالأمر الحتمي. صحيح أنه لا يجدر بك أن تبتاع تذكرة اليانصيب، لكنك قد تربح مالًا لو فعلت. نعم تساعدنا الاحتمالات على الإيقاع بالغشاشين والمجرمين، لكنها إذا استُخدمت بشكل خطأ قد تزج بأبرياء في غياهب السجون. ولهذا كتبنا الفصل السادس.

الفصل الخامس والنصف

معضلة مونتي هول

«معضلة مونتي هول» هي معضلة شهيرة قائمة على الاحتمالات يواجهها المشتركون في برنامج المسابقات «لنعقد صفقة»، الذي بدأ عرضه في الولايات المتحدة عام ١٩٦٣ وما زال يعرض في بعض المحطات في دول كثيرة من العالم. (أتذكر أنني كنت أشاهد هذا البرنامج كلما أصابني المرض وأقعدني عن الذهاب إلى المدرسة الابتدائية). لقد قَدَّم هذا البرنامج هدية للإحصائيين ذكرناها في مقدمة هذا الكتاب. في نهاية كل حلقة كان المتسابق يُدعى ليقف مع مقدم البرنامج أمام ثلاثة أبواب كبيرة: الباب ١ والباب ٢ والباب ٣. يبين مونتي للمتسابق أن هناك جائزة قيمة تتوارى وراء واحد من هذه الأبواب الثلاثة، وهناك عنزة وراء كل واحد من البابين الآخرين. على المتسابق أن يختار بابًا من الثلاثة ويحظى بما وراءه أيًّا كان. (لا أعلم إذا كان يتاح للمتسابق أن يحتفظ بالعنزة لكنني أفترض أن المتسابقين جميعًا كانوا يرحبون بامتلاك السيارة الجديدة).

احتمالات الفوز هنا واضحة مباشرة جدًّا. هناك عنزتان وسيارة. فعندما يقف المتسابق أو تقف المتسابقة بجوار مونتي، يكون أمامه أو أمامها فرصة واحدة من ثلاث فرص لأن ينتقوا ذلك الباب الذي تختفي وراءه السيارة. لكن كما ذكرنا سابقًا، هناك حيلة في هذا البرنامج هي التي خلدت هذا البرنامج ومقدمه في أدبيات الاحتمالات. فبعد أن يختار المتسابق بابًا من الأبواب، يقوم مونتي بفتح باب من البابين اللذين لم يخترهما المتسابق، ودائمًا ما تظهر العنزة خلفه. هنا يسأل مونتي المتسابق عما إذا كان يرغب في تغيير اختياره وأن يتحول عن الباب المغلق الذي اختاره إلى الباب الآخر الذي لم يختره.

لنفترض أن المتسابق اختار الباب ١ في البداية، ففتح مونتي الباب ٣ فظهرت العنزة على المسرح. ما زال هناك بابان مغلقان رقم ١ ورقم ٢. إذا كانت الجائزة القيمة وراء الباب ١ حينها سيفوز المتسابق، أما إذا كانت وراء الباب ٢ فسوف يخسر. هنا يواجه مونتي المتسابق ويسأله عما إذا كان يرغب في تغيير اختياره من الباب ١ إلى الباب ٢، في هذه الحالة عليك أن تتذكر بأن كلا البابين ما زال مغلقًا. المعلومة الجديدة الوحيدة التي تلقاها المتسابق هنا هي أن عنزة ظهرت خلف واحد من البابين اللذين لم يخترهما.

أيجدر به أن يغير قراره؟

نعم. فللمتسابق فرصة تبلغ ⅓ للربح إذا ما تمسك باختياره الأول، أما إذا ما غيَّر قراره تصير فرصته ⅔. إذا لم تصدقني فلتواصل القراءة.

أقر بأن هذه الإجابة قد لا تبدو للوهلة الأولى منطقية بشكل كامل. فظاهريًّا للمتسابق ثلث فرصة للظفر بالجائزة القيمة مهما فعل. هي ثلاثة أبواب مغلقة ولكلٍّ منها فرصة واحدة لأن تكون الجائزة خلفه. ما الفارق إذا تحول المتسابق عن باب مغلق إلى باب مغلق آخر؟

إجابة هذا السؤال تكمن في حقيقة أن مونتي هول يعرف ما يتوارى خلف كل باب من الأبواب الثلاثة. إذا اختار المتسابق الباب رقم ١ وكانت السيارة وراءه، هنا سيفتح مونتي أيًّا من البابين ٢ أو ٣ كي يظهر العنزة.

إذا اختار المتسابق الباب رقم ١ وكانت السيارة خلف الباب رقم ٣، حينها سيفتح مونتي الباب رقم ٢.

عندما يغير المتسابق قراره بعد فتح الباب يستفيد لأنه اختار بابين لا بابًا واحد. سوف أحاول إقناعك بصحة تلك الفكرة عن طريق ثلاث طرق مختلفة.

الطريقة الأولى هي طريقة تجريبية. في عام ٢٠٠٨ كتب جون تيرني كاتب المقالات العمودية في صحيفة النيويورك تايمز عن ظاهرة مونتي هول[١] قامت صحيفة التايمز بإنشاء صفحة إلكترونية تفاعلية تسمح لك أن تلعب تلك اللعبة بنفسك وحتى أن تختار أن تغير قرارك أو لا تفعل. (حتى إنها وضعت بعض السيارات والعنزات الصغيرة التي تطل من وراء الأبواب). تقوم اللعبة برصد مرات نجاحك عندما تقوم بتغيير قرارك في المرة الثانية مقارنة بما يحدث عندما لا تغيره. فلتجربها بنفسك*. لقد دفعت نقودًا لواحدة من بناتي كي تلعبها مائة مرة مبدلة الأبواب في كل مرة. ودفعت نقودًا لأخيها كي يلعبها مائة مرة دون أن يبدل الأبواب. وكانت النتيجة أن ربحت هي ٧٢ مرة وربح هو ٣٣ مرة. لقد تلقى كلٌّ منهما دولارين نظير مجهوده.

والبيانات التي حصلنا عليها من برنامج «لنعقد صفقة» تدعم تلك النتائج. فقد ذكر ليونارد مولدينوف مؤلف كتاب The Drunkard's Walk أن من ربح من أولئك المتسابقين الذين بدلوا اختياراتهم كانوا ضعف من ربحوا ممن لم يفعلوا[٢]

الطريقة الثانية التي سأشرح بها الأمر مبنية على المنطق. لنفترض أن قواعد اللعبة قد عدلت قليلًا، بحيث يبدأ المتسابق باختيار باب من الثلاثة كالمعتاد، لكن بعدها، وقبل

* يمكنك أن تلعب هذه اللعبة على موقع: science/08monty.html?_/08/04/http://www.nytimes.com/2008 .r=2&oref=slogin

أن يفتح أي باب كاشفًا العنزة يقول مونتي: «هل ترغب في أن تتخلى عن اختيارك هذا في مقابل ما وراء البابين الآخرين؟» بمعنى أنك إذا اخترت الباب ١ يمكنك بعدها أن تتخلى عن هذا الاختيار في مقابل أن تحصل على ما يختفي خلف البابين ٢ و٣. وإذا اخترت الباب ٣ في البداية فيمكنك أن تتحول عنه إلى البابين ١ و٢، وهلم جرًّا.

حينها لن يجد المتسابق الاختيار صعبًا. فمن المنطقي أن تترك ذاك الباب الواحد وتختار البابين الآخرين معًا، فهذا يرفع فرص ربحك من ⅓ إلى ⅔. حسنًا إليك الجزء المثير للاهتمام: ذلك هو بالضبط ما يسمح لك مونتي هول بأن تفعله في اللعبة الحقيقية قبل أن يكشف عن العنزة. الفكرة الأساسية هنا أنك إذا اخترت بابين فلا بد أن يكون وراء أحدهما عنزة على أي حال. وعندما يفتح لك بابًا ليكشف العنزة وراءه فهو حينها يسدي لك صنيعًا عظيمًا كما لو كان يقول «إليك ثلثي فرصة لأن تكون السيارة وراء باب من البابين اللذين لم تخترهما، وانظر، هو ليس ذلك الباب».

فلتفكر في الأمر بهذه الطريقة. افترض أنك اخترت الباب ١ حينها يعطيك مونتي الخيار لأن تستبدل به البابين ٢ و٣. إذا ما قبلت العرض متخليًا عن الباب الوحيد ومختارًا البابين، فهذا يعني أنك تتوقع توقعًا منطقيًّا أن تحظى بالسيارة بنسبة ⅔. لكن ماذا إذا فتح مونتي الباب رقم ٣ – واحد من البابين اللذين اخترتهما – كاشفًا العنزة وراءه؟ هل يجدر بك أن تشك في قرارك؟ بالطبع لا. لأن السيارة لو كانت وراء الباب رقم ٣ لكان فتح لك الباب رقم ٢، فهو لم يُرِكَ شيئًا.

عندما تلعب اللعبة بالشكل المعتاد، يعطيك مونتي الاختيار بين الباب الذي اخترته أولًا والبابين الآخرين، واللذان تختفي السيارة وراء واحد منهما فقط. وهو عندما يفتح لك الباب الذي تختفي وراءه العنزة لا يفعل سوى أن يخدمك بإظهار أي البابين الباقيين لا تتواجد السيارة وراءه. تتساوى احتمالات ربحك في كلا السيناريوهين التاليين:

١- اختيار الباب رقم ١ ثم الموافقة على التحول للبابين ٢ و٣ قبل أن يفتح أي باب.

٢- اختيار الباب رقم ١ ثم الموافقة على التحول إلى الباب رقم ٢ بعد أن يكشف مونتي العنزة وراء الباب ٣ (أو اختيار الباب ٣ بعد أن يكشف عن العنزة وراء الباب ٢).

في كلتا الحالتين يعطيك تبديلُ الاختيار فائدة اختيار بابين لا باب واحد، ومن ثم تستطيع أن تضاعف فرصة ربحك من ⅓ إلى ⅔.

أما الطريقة الثالثة التي سأشرح بها الأمر فهي الأكثر تطرفًا من حيث منطقيتها الأساسية. لنفترض أن مونتي هول خيرك بين ١٠٠ باب لا ثلاثة فقط. وبعد أن تختار الباب ٤٧ مثلًا يفتح هو ٩٨ بابًا آخر فتبدو عنزة وراء كل واحد منهم. الآن هناك بابان فقط ظلا مغلقين، الباب رقم ٤٧ (اختيارك الأول) وباب آخر لنفترض أنه رقم ٦١. هل يجدر بك أن تغير قرارك؟

بالطبع عليك أن تفعل. هناك فرصة تبلغ ٩٩ بالمائة أن السيارة تقبع وراء واحد من الأبواب التي لم تخترها أولًا. لقد أسدى لك مونتي معروفًا بأن فتح لك ٩٨ بابًا لم تخترها أنت، وهو يعرف أن السيارة ليست وراء أيٍّ منها. هناك احتمال لا يتجاوز ١ من ١٠٠ أن اختيارك الأول (الباب رقم ٤٧) كان صحيحًا. وهناك احتمال ٩٩ من ١٠٠ أن اختيارك الأول لم يكن صحيحًا. وإذا لم يكن اختيارك الأول صحيحًا فإن السيارة تقبع وراء الباب الآخر رقم ٦١. إذا أردت أن تربح ٩٩ مرة من ١٠٠ فعليك أن تتحول إلى رقم ٦١

للاختصار نقول إنك إذا صرت ذات مرة متسابقًا في برنامج «لنعقد صفقة» فعليك بكل تأكيد أن تبدل اختيارك عندما يقوم مونتي هول (أو بديله) بمنحك هذا الخيار. أما الدرس الأكبر هو أن حدسك الأولي أو الاحتمالات يمكن أن تتيه بك في الضلال.

الفصل السادس

مشكلات الاحتمالات

كيف أوشك مهاويس الرياضيات المبالغون في الثقة بأنفسهم أن يدمروا النظام المالي العالمي

ليست الإحصائيات بأذكى ممن يستخدمونها، بل إنها في بعض الأحيان تجعل الأذكياء يتصرفون بخرق. واحد من أكثر التصرفات غير المسئولة فيما يتعلق باستخدام الإحصائيات، والتي حدثت في الماضي القريب، كان تقديرَ المخاطر في وول ستريت والذي حدث قبيل الأزمة المالية عام ٢٠٠٨. في ذلك الوقت اعتمدت جميع المؤسسات العاملة في الحقل المالي مقياسًا للمخاطر أسموه نموذج القيمة المعرضة للخطر Value at Risk Model، أو VaR اختصارًا. من الناحية النظرية جمع ذلك النموذج بين أناقة المؤشرات (حيث كان يقوم باختزال الكثير من المعلومات إلى رقم مفرد)، وبين قوة الاحتمالات (إذ كان يربط بين الأرباح والخسائر المتوقعة وبين كل أصل من أصول الشركة أو مستوياتها التجارية). يفترض النموذج وجود نطاق كامل من النواتج الممكنة لكل واحد من استثمارات المؤسسة. فمثلًا إذا كانت الشركة تمتلك مجموعة أسهم لشركة جينيرال إلكتريك، فإن هذه الأسهم قد ترتفع أو تنخفض. إذا ما تم حساب النموذج لمدًى زمني قصير، كأسبوع واحد مثلًا، فغالبًا ما سيكون ناتج حساب الأسهم في أول الأسبوع مماثلًا لنظيره في آخره تقريبًا. وهناك احتمال أقل أن تنخفض الأسهم بنسبة ١٠ بالمائة. واحتمال أقل من سابقه أن ترتفع بنسبة ٢٥ بالمائة، وهلم جرًّا.

استنادًا إلى البيانات السابقة المتعلقة بحركات السوق، يستطيع خبراء الكم في الشركة (والذين يُسمَّون «أهل الكم» في ذلك المجال ويسمون «الأثرياء المهاويس» في بقية الأماكن) أن يعيِّنوا قدرًا من المال بالدولارات، ١٣ مليون دولار مثلًا، والذي يمثل أقصى قدر من المال الذي قد تخسره الشركة خلال تلك الفترة الزمنية، ويضعوا عليه احتمالات تبلغ ٩٩ بالمائة. بعبارة أخرى نقول إنه في ٩٩ بالمائة من الحالات لن تخسر الشركة أكثر من ١٣ مليونًا في

تلك الصفقة التجارية، وفي حالة واحدة من مائة ستخسر ما يفوق هذا الرقم.

تذكر هذا الجزء لأنه سيصير مهمًّا عما قريب.

قبل أن تقع الأزمة المالية عام ٢٠٠٨ وضعت المؤسسات ثقتها في نموذج الـ VaR لتقدير المخاطر العامة. فإذا ما أتيحت لُمتاجر واحد ٩٢٣ فرصة مفتوحة (أي استثمار من الممكن أن يتحرك ارتفاعًا أو انخفاضًا) كل واحد من تلك الاستثمارات يمكن أن تقيَّم بالشكل الذي أسلفناه مع أسهم جنيرال إلكتريك، من هنا نستطيع حساب المخاطر التي تحدق بمحفظة ذلك المُتاجر. بل إن تلك المعادلة تأخذ في عين الاعتبار الارتباطَ الحادث بين فرص مختلفة. فمثلًا إذا كان العائدان المتوقعان من استثمارين مختلفين يرتبطان أحدهما بالآخر ارتباطًا سلبيًّا، فلا بد أن خسارة أحدهما يعوضها المكسب في العائد الآخر؛ مما يجعل الجمع بين الاستثمارين أقل خطورة من الاكتفاء بواحد منهما فقط. سيكون لدى رئيس مكتب التجارة فكرة عامة عن أن بوب سميث المُتاجر في الأسهم سيحظى بتقييم مخاطر VaR خلال الأربع والعشرين ساعة التالية يقدر بـ ١٩ مليون دولار، وهو ما يعني مرة أخرى احتمالات قدرها ٩٩ بالمائة. أي إن أكثر ما قد يخسره بوب سميث خلال الأربع والعشرين ساعة المقبلة هو ١٩ مليون دولار في ٩٩ حالة من أصل ١٠٠ حالة.

بل والأفضل من ذلك أن إجمالي المخاطر للمؤسسة ككل يمكن حسابه في أي وقت من خلال المضي لخطوة إضافية في نفس هذه العملية الأساسية. صحيح أن ما ينطوي عليه هذا من آليات رياضية هو أمر غاية في التعقيد، وذلك بسبب العدد الهائل من الاستثمارات التي تقوم بها كل شركة بعملات مختلفة وبمستويات متعددة القوة (بمعنى قدر الأموال التي تم اقتراضها للقيام بتلك الاستثمارات)، والتي تُتداول في الأسواق بمستويات مختلفة من السيولة وأمور أخرى كثيرة؛ لكن رغم هذا كله يمتلك مديرو المؤسسة مقياسًا دقيقًا لعظم المخاطرة التي مرت بها المؤسسة في أي وقت من الأوقات. يقول جو نوسيرا الكاتب السابق في صحيفة النيويورك تايمز: «إن القبول الذي يتمتع به نموذج القيمة المعرضة للخطر وشيوع استخدامه بين مَن هم ليسوا بخبراء في كم المخاطر مردُّه إلى أنه يعبر عن الخطر برقم مفرد أو في صورة مبلغ بالدولارات»[١]. في مؤسسة جيه بي مورجان التي اخترعت هذا النموذج ونقحته وصقلته، يعرف ذلك النموذج باسم تقرير الساعة ٤:١٥؛ وذلك لأنه لا بد أن يتواجد على مكاتب كبار المديرين التنفيذيين الساعة ٤:١٥ عصرًا بالضبط، وذلك كل يوم مباشرة عقب إغلاق سوق الأموال الأمريكية.

كان يفترض بهذا النموذج أن يكون شيئًا جيدًا؛ وذلك لأنه، عندما يتعلق الأمر بتقييم المخاطر، يحبذ أن يتوافر المزيد من المعلومات. فرغم كل هذا لا تزال الاحتمالات أداة قوية، ألم تستخدمها شليتز في حساباتها قبل أن تنفق كل تلك الأموال على اختبار التذوق الذي أجرته في استراحة ما بين شوطَي المباراة النهائية في بطولة كرة القدم الوطنية؟

ليس هذا حقيقيًّا بالضرورة. فقد نُعت نموذج VaR بأوصاف على غرار «كارثة محتملة» و«محض زيف»، وأوصاف أخرى عديدة ليس من المناسب إيرادها في كتاب عائلي يتناول الإحصاء مثل هذا الكتاب. وذلك لأن هذا النموذج تحديدًا قد ألقي عليه باللائمة لأنه هو الذي أطلق بداية الأزمة المالية العالمية. فالانتقاد الأساسي الذي يوجَّه لهذا النموذج أن المخاطر التي يقيمها، والمرتبطة بالأسواق المالية، ليس التنبؤ بها سهلًا كما هو الحال في رمي عملة في الهواء أو حتى عمل اختبار تذوق يفاضل بين نوعين من أنواع الجعة. فتلك الدقة الزائفة التي ينطوي عليها هذا النموذج قد خلقت إحساسًا زائفًا بالأمان. لقد كان ذلك النموذج أقرب إلى عداد سرعة معطوب، وهو ما يراه البعض أسوأ من عدم وجود عداد سرعة على الإطلاق. فإذا ما وضعت الكثير من الثقة في عدادك التالف، لن تلتفت حينها للإشارات الأخرى التي تخبرك بأن سرعتك تتجاوز معدلات الأمان. لكن على العكس إذا لم يكن هناك عداد سرعة مطلقًا فسيجبرك هذا على أن تتلفت حولك باحثًا عن إشارات تخبرك بسرعتك الحقيقية.

فإبان عام ٢٠٠٥، بينما كانت تقارير نموذج VaR توضع على مكاتب المديرين في تمام الساعة ٤:١٥ عصرًا كل يوم عمل، كانت وول ستريت تنطلق بسرعة شديدة جدًّا. ولسوء الحظ كانت هناك مشكلتان كبيرتان في تقييم المخاطر الذي قدمه هذا النموذج مختصرًا. تتمثل المشكلة الأولى في كون الاحتمالات التي اعتمدتها هذه النماذج تستند إلى حركات سوقية سابقة، لكن في الأسواق المالية (على خلاف اختبارات تذوق الجعة) لا يشبه المستقبلُ الماضيَ بالضرورة. فلم يكن هناك مسوِّغ علمي لافتراض أن الحركات السوقية التي تمَّت ما بين عامي ١٩٨٠ و٢٠٠٥ تعتبر المؤشر الأفضل على الحركات السوقية التي ستتم فيما يلي عام ٢٠٠٥. لقد كانت تلك التخيلات الزائفة أقرب إلى ما يفعله العسكريون دائمًا حين يفترضون أن الحرب اللاحقة ستشبه الحرب السابقة. خلال تسعينيات القرن العشرين والعشرية الأولى من القرن الحادي والعشرين كانت البنوك التجارية تَستخدم نماذج إقراضية خاصة بالرهون العقارية وضعت احتمالات مقدارها صفر لأن يحدث انهيار شديد لأسعار العقارات[٢] فلم يسبق أن انخفضت أسعار العقارات بهذه الشدة وهذه السرعة كما فعلت أوائل العام ٢٠٠٧. لكن هذا ما حدث، فقد شرح مدير الاحتياطي الفدرالي السابق آلان جرينسبان هذه الحقيقة لأعضاء لجنة من لجان الكونجرس الأمريكي قائلًا: «لكن ذلك الصرح العلمي تداعى في صيف عام ٢٠٠٧ لأن البيانات التي أدخلت في نماذج إدارة المخاطر كانت تستند بشكل عام إلى ما حدث خلال العقدين الماضيين، والتي مثلت فترة انتعاشة كبيرة. وفي رأيي الخاص إن تلك البيانات لو كانت قد استندت بشكل جيد إلى فترات تاريخية خيم عليها التوتر، لصارت متطلبات رءوس الأموال أعلى بكثير ولصار عالم الأموال في وضع أفضل بكثير»[٣]

أما المشكلة الثانية فتتمثل في كون تلك البيانات التي يقدمها ذلك النموذج حتى لو تنبأت بالمخاطر المستقبلية، فإن نسبة الـ ٩٩ بالمائة التي يقدمها نموذج VaR لا جدوى منها، بل وتنطوي على خطورة كذلك على اعتبار أن نسبة الـ ١ بالمائة المتبقية يمكن أن تفسد عليك الأمر كله. يوضح ديفيد إينهورن مدير إحدى المحافظ الوقائية الأمر بقوله: «هذا النموذج هو أقرب إلى وسادة سيارة هوائية تعمل في جميع الأوقات إلا عندما تتعرض لحادث سيارة». فإذا قدَّر النموذج القيمة المعرضة للخطر لمؤسسةٍ ما بكونها ٥٠٠ مليون دولار، فيمكن أن يفهم من هذا أن هناك فرصة قدرها ٩٩ بالمائة أن تخسر الشركة ما لا يتجاوز ٥٠٠ مليون دولار خلال تلك الفترة الزمنية المحددة. لكن، يا قوم، هذا يعني أيضًا أن هناك فرصة ١ بالمائة أن تخسر الشركة ما يتجاوز ٥٠٠ مليون دولار بكثير جدًّا في ظل ظروف معينة. بل الواقع أن تلك النماذج لم تتكلم بتاتًا عن نسبة الـ ١ بالمائة تلك وعما سينتج عنها. فلم يكرِّس سوى القليل من الاهتمام إلى «المخاطرات الذيلية» أو المخاطرات الصغيرة (سميت نسبةً إلى ذيل التوزيع) لحدوث نواتج كارثية. (فمثلًا إذا قدت سيارتك من الحانة إلى المنزل ومستوى الكحول في دمك ٠,١٥ فهناك احتمالات قدرها يقل عن ١ بالمائة أن حادثة ستقع لك وتموت، لكن هذا لا يعني أن ذلك أمر لك أن تفعله بشكل منطقي). وهكذا فقد قامت العديد من الشركات بارتكاب ذلك الخطأ من خلال وضع افتراضات غير واقعية عن جاهزيتها للأحداث النادرة الوقوع. وضح هانك بولسون وزير الخزانة السابق الأمر بقوله إن الكثير من الشركات قد افترضت أنها تستطيع أن تجمع النقد في لمحة عين من خلال بيع بعض الأصول[٤]. لكن خلال الأزمات تحتاج كل الشركات للنقد، ولذلك فكلها تحاول بيع نفس نوع الأصول. إن هذا هو المكافئ المتعلق بتقييم المخاطر لمقولة: «إنني لا أحتاج لأن أخزن الماء، فإذا ما حدثت كارثة طبيعية كل ما عليَّ أن أفعله هو أن أذهب للمتجر وأبتاع بعضه». لكن بالطبع بعد أن يضرب كويكب بلدتك سيُهرَع خمسون ألف شخص غيرك إلى المتجر كي يشتروا الماء، وعندما تصل أنت إليه ستجد نوافذه مهشمة وأرففه خاوية.

إن كونك لم تفكر في إمكانية أن يضرب كويكب هائل بلدتك فيسويها بالأرض هو بالضبط المشكلة في نموذج VaR. مرة أخرى نعود لجو نوسيرا كاتب المقالات في صحيفة النيويورك تايمز، إذ يقوم بتلخيص أفكار نيكولاس تاليب مؤلف كتاب The Black Swan: The Impact of the Highly Improbable، ويكيل النقد اللاذع لنموذج VaR قائلًا: «إن أكبر المخاطر ليست هي التي تستطيع رؤيتها وتقديرها، وإنما هي تلك التي لا تستطيع أن تراها، وبالتالي لا تستطيع أبدًا أن تقدرها. تلك المخاطر التي تبدو بعيدة عن حدود الاحتمالات الطبيعية إلى درجة أنك لا تتخيل أنها قد تحدث خلال حياتك رغم أنها تحدث بالطبع أكثر مما تظن».

إن كارثة ذلك النموذج هي، بشكل أو بآخر، نقيض مثال جعة شليتز الذي أوردناه في الفصل الخامس؛ حيث كانت شليتز تعمل على توزيع احتمالات معروف. فأيًّا كانت البيانات التي حازتها الشركة عن أرجحية اختيار المتذوقين معصوبي الأعين لجعة شليتز، فقد كانت تقديرًا جيدًا لما سيفعله أولئك المتذوقون على الهواء مباشرة خلال استراحة ما بين الشوطين. بل إن شليتز تعاملت مع الجانب السلبي في الموضوع عن طريق إجراء الاختبار كاملًا على رجال قالوا إنهم يفضلون أنواعًا أخرى من الجعة. فحتى إذا لم يختر أكثر من خمسة وعشرين شاربًا لجعة ميشيلوب جعة شليتز (وهو ناتج متدنٍّ جدًّا بشكل يكاد يكون مستحيلًا) فستظل شليتز قادرة على أن تقول إن واحدًا من كل أربعة شاربين للجعة سوف يغيرون مشروبهم المفضل. لكن ربما كان أهم ما في الموضوع أن ذلك كان يتعلق بنوع من أنواع الجعة لا بالنظام المالي العالمي. لقد ارتكب خبراء وول ستريت ثلاثة أخطاء أساسية. الأول أنهم خلطوا بين الدقة والتحديد؛ فقد كان نموذج VaR أشبه بمحدد نطاق الجولف الذي امتلكته، والذي كان مهيَّأً للقياس بالأمتار لا الياردات، أي إنه كان دقيقًا لكنه كان خطأً. كان الخطأ في التحديد هو الذي قاد مديري وول ستريت لأن يعتقدوا بأنهم يسيطرون على المخاطر، في حين أن هذا لم يكن صحيحًا. والثاني أن تقديرات الاحتمالات كانت خطأ، فكما أشار آلان جرينسبان في شهادته التي أوردناها في جزء سابق من هذا الفصل أن سنوات الهدوء والرخاء النسبي التي سبقت عام ٢٠٠٥ لم يكن ينبغي أن تستخدم لإنشاء توزيع احتمالي لما قد يحدث في الأسواق المالية خلال الأعوام التالية. فهذا التفكير يشبه أن تدلف أنت إلى صالة قمار متوقعًا أن تربح في لعبة الروليت بنسبة ٦٢ بالمائة لأن ذلك هو ما حدث آخر مرة قامرت فيها. ستكون ليلة طويلة باهظة التكلفة. أما الخطأ الثالث فتمثل في تجاهل «المخاطر الذيلية». فنماذج VaR تنبأت بما سيحدث في ٩٩ مرة من أصل ١٠٠. هذه هي الطريقة التي تعمل بها الاحتمالات (وهو الأمر الذي سيؤكد عليه القسم الثاني من الكتاب مرارًا وتكرارًا). فغير المتوقع من الأمور يحدث. والصواعق لا تفتأ تضرب الناس، وسبق أن أسقطت أمي كرة الجولف في الفتحة من ضربة واحدة وكررتها مرتين أُخريين.

لقد كانت الغطرسة التي سادت البنوك التجارية وَوول ستريت هي السبب الأخير الذي ساهم في حدوث أفدح انكماش مالي عالمي منذ الكساد الكبير. لقد أدت هذه الأزمة التي بدأت عام ٢٠٠٨ إلى خسارة ثروات تقدَّر بتريليونات الدولارات في الولايات المتحدة، ورفعت معدلات البطالة بنسبة تجاوزت الـ ١٠ بالمائة، وخلقت موجات من الحجز على المساكن المرهونة وإخفاقات في الشركات، وكبَّلت الكثير من الحكومات في أنحاء العالم بالكثير من الديون التي حاولوا بها أن يحتووا ذلك الخراب الاقتصادي. كانت تلك مفارقة مؤسفة؛ نظرًا لأن ذلك النموذج المنمق كان يعد أداة مصممة لتقليل المخاطر.

إن الاحتمالات تقدِّم لنا مجموعة من الأدوات النافعة الفعالة يمكن أن يسخَّر الكثير منها بشكل صحيح بغرض فهم العالم الذي يحيطنا، ويمكن أن يُستخدم بشكل غير صحيح فيُشيع الفوضى في أرجائه. ولأنني متمسك بتشبيه «الإحصائيات بالأسلحة الفتاكة»، الذي لم أتوقف عن استخدامه عبر صفحات الكتاب، فسوف أعيد صياغة مقولة لجماعة الضغط التي تنادي بحق حمل السلاح: إن الاحتمالات لا ترتكب الأخطاء، وإنما يرتكبها من يستخدمون الاحتمالات. سوف يمضي هذا الفصل في بيان الشائع من الأخطاء والمعضلات الأخلاقية والأمور التي فُهمت بشكل خطأ والمرتبطة بالاحتمالات.

افتراض كون بعض الأحداث منفصلة في حين أن هذا غير صحيح. تبلغ احتمالات رمي العملة فتسقط كاشفة وجه الرأس ½. أما احتمالات حدوث ذلك مرتين متتاليتين فتبلغ ٢(½) أو ¼؛ حيث إن أرجحية وقوع حدثين منفصلين هي نتاج احتمالاتهما منفصلين. والآن بعد أن تسلحت بتلك المعرفة، دعنا نفترض أنك عُيِّنتَ رئيس فريق تقييم المخاطر في إحدى كُبرَيات شركات الطيران. يخبرك مساعدك أن احتمالات حدوث عطل في محرك إحدى الطائرات، لأي سبب من الأسباب خلال إحدى الرحلات العابرة للمحيط الأطلسي، تبلغ ١ من كل ١٠٠٠٠٠ رحلة. بالنظر إلى عدد الرحلات التي تعبر الأطلسي، لا يمكن أن تقبل تلك المخاطرة. لكن معظم الطائرات التي تعبر الأطلسي، لحسن الحظ، ذات محركين على الأقل. قام مساعدك بحساب أن خطورة تلف المحركين معًا فوق المحيط الأطلسي لا بد أن تكون ٢ (١/ ١٠٠٠٠٠) أو ١ إلى ١٠ مليار، وهو الأمر الذي يعتبر مخاطرة معقولة. حينها يجدر بك أن تطلب من مساعدك أن يذهب ويستخدم أيام عطلاته المتبقية قبل أن يفصل من عمله. فحدوث تلف بكلا المحركين لا يُعتبر حدثين منفصلين. فإذا اخترقت الطائرة سربًا من الإوز أثناء إقلاعها فإن الأرجح أن يتأذى المحركان بذات الدرجة. وهو الأمر الذي ينطبق أيضًا على عوامل كثيرة تؤثر على أداء المحرك، كحالة الطقس أو الإهمال في الصيانة. فإذا ما تلف واحد من المحركين، فإن احتمالات تلف المحرك الثاني تبلغ ما هو أكثر بكثير من ١ إلى ١٠٠٠٠٠.

هل اتضح الأمر؟ لكنه لم يكن واضحًا خلال تسعينيات القرن العشرين، وقت أن ارتكب ممثلو الادعاء البريطانيون سقطة قضائية كبيرة بسبب استخدام غير صحيح للاحتمالات. فكما صار في مثال تلف محركات الطائرة الافتراضي ذلك، تَمثَّل الخطأ الإحصائي في افتراض كون مجموعة من الأحداث منفصلة أحدها عن الآخر (كما في رمي العملة) لا مرتبطة بعضها ببعض (عندما يجعلنا ناتج معين نرجِّح حدوث ناتج مشابه له مستقبلًا). لقد كان أمرًا خطأً بالفعل وتَرتَّب عليه أن زَجَّ في السجن بأبرياء.

كما برز ذلك الخطأ في سياق قضية متلازمة موت الرضع المفاجئ (SIDS)، وهي ظاهرة يموت فيها أطفال رضع أصحاء جدًّا في مهدهم. (يسميها البريطانيون «الموت في

المهد»). صارت تلك المتلازمة لغزًا طبيًّا جذب الانتباه؛ لأنها صارت أكثر أسباب وفيات الرضع شيوعًا*. ولأن وفيات الرضع تلك كانت محيرة، ولأنها فُهمت بشكل خطأ فقد أثارت الريبة. وأحيانًا ما كانت هذه الريبة مبررة. فقد استُخدمت هذه المتلازمة أكثر من مرة للتغطية على جرائم الإهمال الأبوي أو إيذاء الأطفال؛ وذلك لأن فحوصات ما بعد الوفاة لا تميز الوفيات الطبيعية عن تلك التي ترتبت على جرائم؛ لهذا فقد اعتقد ممثلو الادعاء والقضاة البريطانيون أن السبيل لتمييز الوفيات الطبيعية عن الجرائم هو التركيز على الأسر التي تتكرر فيها حالات «الموت في المهد». كان السير روي ميدو طبيب الأطفال البريطاني البارز شاهدًا خبيرًا استعانت به المحاكم كثيرًا في هذه القضايا. عن هذا قالت مجلة الإيكونوميست الإخبارية البريطانية: «إن ما عُرف باسم قانون ميدو – فكرة أن وفاة رضيع واحد هي مأساة ووفاة رضيعين هي أمر مثير للريبة، أما وفاة ثلاثة فهو جريمة قتل – قائم على مفهوم أنه إذا ندر وقوع حدث ما، فإن تكراره مرتين أو أكثر في ذات الأسرة هو أمر يستبعد حدوثه عن طريق الصدفة»[٥] بيَّن السير ميدو لهيئة المحلفين أن احتمال حدوث حالتَي وفاة طبيعيتين لرضيعين من نفس الأسرة هي ١ من ٧٣ مليونًا. ثم شرح حساباته على النحو التالي: بما أن حدث الوفاة في المهد نادر فاحتماله ١ من ٨٥٠٠، إذن فاحتمالات حدوث حالتَي وفاة بشكل طبيعي هي ٢(١/ ٨٥٠٠)، وهو ما يساوي ١ من ٧٣ مليونًا تقريبًا. وهو الرقم الذي تفوح منه رائحة الجريمة. ولقد كان هذا ما أقرَّته هيئة المحلفين وأرسلت الكثير من الآباء إلى السجن استنادًا إلى هذه الشهادة عن إحصائيات الوفاة في المهد (والتي كانت عادة تقدم بلا أدلة طبية داعمة دالة على إهمال أو إيذاء). بل إنه في بعض الأحيان كان الأطفال يبعدون عن ذويهم عقب الولادة بسبب وفيات أشقائهم غير المفسرة.

وضحت مجلة الإيكونوميست كيف كان سوء فهم انفصال الأحداث إحصائيًّا هو ما عاب شهادة ميدو:

> هناك عيب واضح في هذا المنطق، كما أشارت الجمعية الإحصائية الملكية. إن حسابات الاحتمالات لا يعيبها شيء، هذا إذا تيقَّنَّا من أن وفيات المهد عشوائية تمامًا ولا يربط بينها أي عامل مجهول. لكن مع أمر ملغز كَوَفيات المهد، فمن الممكن جدًّا أن يكون ثمة رابط – كرابط وراثي على سبيل المثال، وهو الأمر الذي لا يجعل تكرار تلك الوفيات المأساوية أمرًا مستبعدًا. ولقد ظل العلماء يطرحون فكرة وجود هذا الرابط منذ أن تمَّت إدانة أولئك النسوة.

* لا تزال متلازمة موت الرضع المفاجئ لغزًا طبيًّا رغم تحديد أكثر من عامل متعلق بالمخاطر. فمثلًا يمكن تقليل وفيات الرضع كثيرًا إذا ما أرقدوا على ظهورهم.

في عام ٢٠٠٤ أعلنت الحكومة البريطانية أنها سوف تراجع ٢٥٨ محاكمة أُدين فيها آباء وأمهات بقتل أطفالهم الرضع.

عدم إدراك متى تكون الأحداث منفصلة. نوع آخر من الأخطاء يُرتكب عندما تكون الأحداث منفصلة لكنها لا تعامل على أنها كذلك. إذا وجدتَ نفسك في صالة لعب قمار (وهي مكان لا يجدر بك، من وجهة نظر إحصائية، أن ترتاده) فسوف ترى أناسًا يتطلعون بتَوْقٍ إلى النرد أو إلى أوراق اللعب قائلين «لقد حان وقت الربح». فإذا استقرت كرة الروليت على اللون الأسود خمس مرات متتالية، فلا بد أنها الآن ستستقر على اللون الأحمر. لا، لا، ثم لا؛ فاحتمالات استقرار الكرة على اللون الأحمر لا تزال كما هي لم تتغير: ٣٨/١٦. وإن يظن أحد ما غير ذلك فهو إذن يمر بما يسمى «مغالطة المقامر». فالواقع أنك إذا ما رميت عملة سليمة مليون مرة فسقطت في المرات المليون كاشفة وجه الرأس فستظل احتمالات انكشاف وجه الذيل في المرة التالية ½. فالتعريف المحدد للانفصال الإحصائي بين حدثين هو أن ناتج أحدهما لا يؤثر على ناتج الآخر. وحتى لو لم يقنعك الإحصاء، فربما تنظر إلى الأمر من ناحية الفيزياء: فكيف لسقوط العملة مرات عديدة متتالية كاشفة وجه الذيل أن يرجِّح كونها ستسقط كاشفة الوجه الآخر في المرة القادمة؟

بل إننا قد نجد مفهوم التكرار المتتالي خادعًا حتى في الرياضة. واحدة من أشهر الدراسات العلمية، التي تتناول الاحتمالات، وأكثرها إثارة للاهتمام قامت بتفنيد ذاك المفهوم الذي شاع عن أن لاعبي كرة السلة يقومون بشكل دوري برميات ممتازة متتالية خلال مباراة بعينها. بالطبع سوف يقول لك مشجعو الرياضات إن اللاعب الذي أحرز هدفًا لتوِّه أقرب لإحراز هدف ثانٍ من الآخر الذي أخطأ الرمية فيما يعرف بنظرية «اليد الساخنة». لكن هذا لا يتفق مع البحث الذي قام به توماس جيلوفيتش وروبرت فالون وعاموس تفيرسكي، الذين درسوا تلك اليد الساخنة بثلاث طرق مختلفة[٦] فقد قاموا بتحليل بيانات الرميات التي رماها فريق فيلادلفيا سيكستي سيفنرز خلال المباريات التي لعبها خلال موسمَي ١٩٨٠ و١٩٨١. (في ذلك الوقت لم تكن بيانات بقية فرق الرابطة الوطنية لكرة السلة متاحة). فلم يجدوا أي ارتباط إيجابي بين نواتج الرميات المتتالية. ثم قاموا بدراسة نفس البيانات المتعلقة بالرميات الحرة التي رماها فريق بوسطن سيلتيكس، فخرجوا بنفس النتائج. وأخيرًا قاموا بعمل تجربة مضبوطة على أعضاء فريقَي الرجال والنساء لكرة السلة بجامعة كورنيل. فقد أحرز اللاعبون ما متوسط نسبته ٤٨ بالمائة من أهدافهم الميدانية بعد إحراز هدف قبله، وأحرزوا ما نسبته ٤٧ بالمائة عقب إضاعتهم لهدف قبله. ولقد جاء الارتباط سلبيًّا لدى أربعة عشر لاعبًا من أصل ستة وعشرين لاعبًا بين إحراز هدف وإحراز هدف يليه. لاعب واحد فقط هو الذي ظهر لديه ارتباط إيجابي كبير بين الرمية والتي تليها.

لكن معظم مشجعي كرة السلة سوف يخبرونك بشيء آخر. فمثلًا ذَكر ما نسبته ٩١ بالمائة من مشجعي كرة السلة الذين استُطلعت آراؤهم بجامعتَي ستانفورد وكورنيل أن اللاعب بعد أن يحرز هدفين أو ثلاثة، تزيد فرصة إحرازه لهدف آخر عنها إذا ما كان قد أهدر الفرصتين السابقتين. تتمثل أهمية دراسة «اليد الساخنة» تلك في الفرق بين إدراك الظاهرة والحقيقة التجريبية. فقد أورد الكتَّاب الثلاثة ملحوظة في بحثهم تقول: «إن مفاهيم الناس البديهية عن العشوائية تختلف تمامًا عن قوانين الاحتمالات». فإننا نرى أنماطًا نظامية ليست موجودة من الناحية الواقعية.

كما في العناقيد السرطانية.

العناقيد تتكون. لعلك قد قرأت خبرًا في صحيفة ما أو ربما شاهدت تقريرًا في نشرة إخبارية ما يروي عن أن عددًا غير معتاد من الناحية الإحصائية من البشر القاطنين بمنطقة بعينها قد أصيبوا بنوع نادر من أنواع السرطان. لا بد أن السبب يكمن في مياه الشرب أو محطة الطاقة المحلية أو لعله برج تقوية شبكة الهاتف المحمول. بالطبع من الممكن أن يكون أيٌّ من هذه الأشياء هو ما سبب تلك المضارِّ الصحية غير المألوفة. (في فصول لاحقة سوف نستكشف كيف يمكن للإحصاء أن يحدد تلك العلاقات السببية). لكن هذا العنقود من الحالات قد يكون أيضًا محض مصادفة، حتى عندما يبدو عدد الحالات غير معتاد أبدًا. نعم تبلغ احتمالات إصابة خمسة أشخاص في نفس المدرسة أو نفس الكنيسة أو نفس محل العمل بنوع نادر من سرطان الدم واحدًا إلى مليون، لكن هناك الملايين من المدارس والكنائس ومحال العمل. وليس من المستبعد جدًّا أن يصاب خمسة أشخاص في واحد من هذه الأماكن بسرطان الدم النادر ذاك. لكننا لم نفكر في كل المدارس والكنائس ومحال العمل التي لم تحدث بها تلك الإصابات. كي نستخدم تنوعًا آخر على نفس المثال الأول، نقول إن احتمال فوز أحد الأشخاص باليانصيب قد تكون ١ من ٢٠ مليون، لكن رغم هذا لا نندهش عندما يفوز أحدهم لأن التذاكر التي تباع تبلغ أعدادها الملايين. (رغم نفوري من اليانصيب بشكل عام لا بد أن أعترف بإعجابي بشعار يانصيب إيلينوي: «أحدهم سيربح اليانصيب، ولا مانع أن تكون أنت»).

إليك تمرين أجريه مع طلبتي كي أثبت نفس الفكرة. وكلما كان عدد طلبة الصف أكبر، جرى التمرين بشكل أفضل. في التمرين أطلب من كل مَن في الصف أن يمسك بعملة ثم يقف. ثم نقوم كلنا برمي العملة، وكل من تسقط عملته كاشفة وجه الرأس عليه أن يجلس. إذا افترضنا أن من في الصف كانوا ١٠٠ طالب، فسوف يجلس نحو ٥٠ طالبًا عقب الرمية الأولى. ثم نكرر العملية مرة ثانية فيظل ٢٥ طالبًا أو نحوهم واقفين، وهلم جرًّا. وكان من المعتاد أن يظل طالب واحد واقف بعد أن كشف وجه الذيل خمس مرات متتالية أو ستة. عند هذه النقطة أسأل ذلك الطالب سؤالًا على غرار: «كيف فعلتها؟»، أو «كيف

تتدرب على رمي العملة فتسقط كاشفة وجه الذيل مرات عديدة متتالية بهذه الكفاءة؟»، أو«هل تتبع حمية غذائية خاصة ساعدتك على تحقيق هذا الإنجاز المذهل؟». دائمًا ما تستثير هذه الأسئلة ضحكات الطلبة؛ وذلك لأنهم شاهدوا العملية كاملة تتكشف أمامهم وهم يعرفون أن من أصابوا وجه الذيل ست مرات متتالية لا يتمتعون بأي مهارات خاصة في رمي العملات. بل إنه، أو إنها، قد تصادف أنه رمى العملة فكشفت الكثير من وجوه الذيل. لكننا عندما نشهد حدثًا شاذًّا خارجًا عن السياق كهذا، دائمًا ما نفترض وجود أمر آخر مسئول عن حدوثه بخلاف العشوائية.

مغالطة ممثلي الادعاء. لنفترض أنك استمعت إلى شهادة في إحدى المحاكم جاءت على غرار ما يلي: (١) إن عينة الحمض النووي التي وُجدت في مسرح الجريمة تطابق العينة التي أخذت من المدعى عليه، و(٢) هناك فرصة تبلغ واحدًا في المليون أنْ تُطابِقَ العينةُ التي وُجدت في مسرح الجريمة أيَّ شخص بخلاف المدعى عليه (لأجل هذا المثال لنفترض أنك تستطيع افتراض صحة احتمالات ممثل الادعاء). بناءً على هذا الدليل هل ستصوت لصالح الإدانة؟

بالقطع لا أتمنى ذلك.

تحدث مغالطة الادعاء عندما يتم إهمال السياق المحيط بالدليل الإحصائي. لتلك الحالة سيناريوهان كل منهما يبين استخدام دليل الحمض النووي في الادعاء على المتهم.

المدعى عليه ١: هذا المدعى عليه المحب للضحية، والذي واجه صدًّا في مقابل حبه ذلك، وُجد على بعد ثلاثة أحياء سكنية من مسرح الجريمة حاملًا سلاح الجريمة. وعقب إلقاء القبض عليه أجبرته المحكمة على أن يقدم عينة من حمضه النووي، والتي طابقت عينة من شعرة وجدت في مسرح الجريمة.

المدعى عليه ٢: هذا المدعى عليه سبق وأن أُدين في جريمة مماثلة في ولاية أخرى قبل عدة سنوات. ونتيجة لإدانته تلك فقد وُضعت بصمة حمضه النووي في قاعدة بيانات قومية تحوي ما يزيد على المليون عينة لمرتكبي جرائم العنف. قورنت العينة التي أُخذت من الشعرة التي وجدت في مسرح الجريمة بالعينات الموجودة في قاعدة البيانات القومية، فطابقت ذلك الشخص الذي لم تُكتشف صلة بينه وبين الضحية.

كما أشرنا سابقًا، يستطيع ممثل الادعاء في كلتا الحالتين أن يقول وهو محق إن عينة الحمض النووي المأخوذة من مسرح الجريمة تطابق عينة المدعى عليه، وأن هناك فرصة لا تتجاوز واحدًا من المليون أنْ تُطابِقَ عينةً لشخص آخر غيره. لكن في حالة المدعى عليه رقم ٢ هناك فرصة كبيرة جدًّا أن يكون هو ذاته ذلك الشخص العشوائي الآخر، ذلك الاحتمال الذي يبلغ ١ من مليون يفترض أن يطابق حمضه النووي حمض القاتل

الحقيقي؛ وذلك لأن احتمال العثور على مصادفة تبلغ واحدًا من مليون تصير عالية نسبيًّا إذا ما قورنت العينة بمحتويات قاعدة بيانات بها عينات مليون شخص.

الارتداد نحو المتوسط (أو الانحدار نحو المتوسط). لعلك قد سمعت عن لعنة مجلة سبورتس إلاستريتد، والتي تتمثل في كون أي فريق أو أي لاعب فردي توضع صورته على غلاف تلك المجلة لا بد وأن يشهد مستواه انحدارًا بعدها. أحد تفسيرات ذلك الأمر هو أن الظهور على غلاف المجلة له تأثير عكسي على الأداء اللاحق. أما التفسير الأكثر منطقية من الناحية الإحصائية فهو أن هذه الفرق وهؤلاء الرياضيون لا يظهرون على غلاف المجلة إلا بعد أن يحققوا إنجازات جيدة غير معتادة (كأن يربحوا عشرين مباراة متتالية)، ومن ثم لا يكون تراجع أدائهم اللاحق إلا عودة للوضع الطبيعي أو المتوسط. تعرف هذه الظاهرة باسم الارتداد إلى المتوسط. تخبرنا هذه الظاهرة بأن أي قيم متطرفة – أي ملاحظة تقع بعيدة جدًّا عن المتوسط في أي من الاتجاهين – عادة ما يتبعها ناتج أكثر توافقًا مع المتوسط الطويل المدى.

الارتداد إلى المتوسط يمكنه أن يفسر لماذا يدفع نادي شيكاجو كابس رواتب خيالية لِلَاعبين غير مقيدين لا يفتئون يحبطون المشجعين أمثالي بعد التعاقد معهم. فأولئك اللاعبون يتسنى لهم أن يفاوضوا النادي على رواتب ضخمة عقب موسم استثنائي أو اثنين. فارتداء زي شيكاجو كابس لا يجعل أداء أولئك اللاعبين سيئًا بالضرورة (رغم أنني لا أستطيع أن أستبعد ذلك الاحتمال كلية)، وإنما يدفع النادي هذه الأموال الكثيرة لهؤلاء النجوم إبان نهاية فترة استثنائية – قيمة متطرفة تستمر لعام أو اثنين – وبعدها يرتد أداؤهم إلى ما يقارب المتوسط الطبيعي.

ونفس الظاهرة يمكن أن تفسِّر السبب في أن التلاميذ الذين يُبلون بلاءً حسنًا على خلاف عادتهم في أحد الاختبارات، يتراجع مستواهم نوعًا ما مرة أخرى عند إعادة اختبارهم، بينما يعود أداء التلاميذ الذين يؤدون الاختبارات بشكل سيء على عكس المعتاد إلى المتوسط الطبيعي عند إعادة اختبارهم. واحد من تفسيرات الارتداد نحو المتوسط ذلك هو أن الأداء – العقلي والجسدي كليهما – يتألفان من بعض الجهد الذي ينطوي على موهبة، إضافة إلى عامل الحظ جيدًا كان أو عاثرًا. (يسمي الإحصائيون هذا بالخطأ العشوائي). ففي أي حالة يؤدي فيها أولئك الذين يحققون ما هو أعلى من المتوسط بكثير خلال فترة معينة، لا بد أن الحظ يحالفهم، والعكس ينطبق على من يؤدون أداءً يدنو كثيرًا عن المتوسط، فلا بد أن الحظ يُجانبهم. (في حالة الاختبار، تخيل التلاميذ الذين يختارون بين الإجابات الصحيحة والخطأ، أما في حالة لاعبي البيسبول تخيل رمية قد تُحسب خطأً أو قد تقع في منطقة الثلاث نقاط). وعندما تنقشع سحابة الحظ الحسن أو الحظ السيء – ولا بد أن تنقشع – فإن الأداء المنتظر سيكون قريبًا من المتوسط.

تخيل أنني أحاول أن أشكل فريقًا من نجوم رمي العملة (وأنا واقع تحت تأثير انطباع زائف بأن الموهبة مطلوبة لرمي العملات في الهواء). وبعد أن لاحظت تلميذًا أصاب وجه الذيل ست مرات متتالية، عرضت عليه عقدًا لمدة عشر سنوات بقيمة ٥٠ مليون دولار. غني عن القول أنني سأصاب بإحباط شديد عندما يصيب ذلك التلميذ وجه الذيل بنسبة ٥٠ بالمائة على مدى السنين العشر.

للوهلة الأولى قد يبدو الارتداد نحو المتوسط يتعارض مع «مغالطة المقامر». فبعد أن يصيب التلميذ وجه الذيل ست مرات متتالية، هل «حان وقت» محاولته إصابة الرأس أم لا؟ احتمالات إصابته وجه الرأس ما زالت كما هي ½. وحقيقة أنه أصاب وجه الذيل مرات كثيرة لا تعني أبدًا أرجحية إصابته لوجه الرأس في المرة التالية. فكل رمية هي حدث منفصل مستقل. لكننا يمكن أن نتوقع توافق نتائج الرميات التالية مع تنبؤات الاحتمالات، والتي تمنح النصف لوجه الرأس والنصف الآخر لوجه الذيل، على خلاف ما حدث في السابق، ألا وهو وجه الذيل على الدوام. عندما لا يصادف اللاعب إلا وجه الذيل يتكون يقين افتراضي أنه سيصيب وجه الرأس خلال العشر أو العشرين أو المائة رمية المقبلة. وكلما زاد عدد الرميات، تشابَه الناتج مع متوسط النواتج المقدر بـــ ٥٠ – ٥٠، كما تنص عليه تنبؤات قانون الأرقام الكبيرة. (وإلا شرعنا في التفتيش عن أدلة على حدوث احتيال).

في ملاحظة جانبية مثيرة للفضول نقول هنا إن بعض الباحثين قد وثَّقوا ظاهرة أخرى متعلقة بمجلة بيزنس ويك. فعندما يحصل مديرون تنفيذيون على جوائز رفيعة، من بينها أن يحوز أيٌّ منهم لقب «المدير الأفضل» من مجلة بيزنس ويك، يتبع هذا مباشرة تراجعٌ في أداء شركاتهم خلال السنين الثلاث التالية، سواء من ناحية حسابات الأرباح أو من ناحية أسعار أسهمها. لكن، وعلى خلاف ما يحدث مع مجلة سبورتس إلاستريتد، يفوق هذا التراجع الارتداد إلى المتوسط. ذكر أورليك ماليمنديير عالم الاقتصاد بجامعة كاليفورنيا ببيركلي، وجيوفري تايت عالم الاقتصاد بجامعة كاليفورنيا بلوس آنجيليس، أن هؤلاء المديرين التنفيذيين بعد أن يحققوا تلك مكانة ويصيروا «نجومًا»، لا يلبثوا أن يتشتت تفكيرهم من جراء شهرتهم تلك[٧]. فيشرعون في كتابة مذكراتهم، ويدعون إلى جلسات لا علاقة لها بأعمالهم، ويأخذون يبحثون عن زوجات أو أزواج جميلي الشكل يتفاخرون بهم أمام الناس (لقد ذكر الكاتبان السببين الأولين وذكرتُ أنا الثالث من عندياتي لأنني أراه سببًا معقولًا كذلك). كتب ماليمنديير وتايت يقولان: «إن النتائج التي توصلنا إليها تشير إلى أن ثقافة نجوم الأعمال تلك التي تولدها وسائل الإعلام تؤدي إلى انحرافات سلوكية تتجاوز الارتداد إلى المتوسط». بعبارة أخرى نقول إنك عندما ترى المدير التنفيذي للشركة التي تملك أسهمها قد ظهر على غلاف مجلة بيزنس ويك، فلتَبِعْ أسهُمك.

التفرقة الإحصائية. متى يسمح لنا بأن نتصرف استنادًا إلى قواعد ما تخبرنا

الاحتمالات بأنه سوف يحدث، ومتى لا يسمح لنا؟ في عام ٢٠٠٣ قامت آنا ديماننتوبولو، المفوضة الأوروبية للعمالة والشئون الاجتماعية، باقتراح إعلان توجيهي يفرض على شركات التأمين ألَّا تفرق في الأقساط بين الرجال والنساء؛ لأن هذا يخرق مبدأ المساواة في المعاملة الذي يتبناه الاتحاد الأوروبي[٨]. لكن شركات التأمين لا ترى في تحديد المبلغ الذي يدفعه العميل بناءً على جنسه تفرقة، وإنما هو بالنسبة إليها محض إحصاء. فالرجال عادة ما يدفعون أكثر للتأمين على سياراتهم لأنهم يقعون في عددٍ أكبر من الحوادث. بينما تدفع النساء مبلغًا أكبر لتأمين المعاش (وهو نوع من أنواع التأمين يدفع للمؤمن عليه مبلغًا ثابتًا من المال حتى وفاته)؛ وذلك لأنهن يعشن لأعمار أطول. من المعروف أن الكثير من النساء يقعن في حوادث سيارات أكثر من الرجال، والكثير من الرجال يعيشون لأعمار أطول من النساء، لكن شركات التأمين، كما ذكرنا في الفصل السابق، لا تبالي بهذا. إنهم لا يبالون إلا بمتوسط ما يحدث لأنهم إن أصابوا في توقُّع ذلك جَنَوا الكثير من الأموال. الأمر المثير للاهتمام فيما قامت به المفوضية الأوروبية من سياسة لمنع التفرقة التي تقوم بها شركات التأمين طبقًا لجنس العميل، وهي السياسة التي بدأ تنفيذها عام ٢٠١٢، هو أن السلطات لا تدَّعي بأن الجنس لا يرتبط بالمخاطر التي يؤمن ضدها، وإنما كل ما تفعل هو أن تخرج بتصريحات مفادها أن تلك التفاوتات القائمة على أساس الجنس هي أمر غير مقبول*.

للوهلة الأولى يبدو هذا الإعلان إيماءة مزعجة تجاه إصلاح سياسي. لكنني بعد أن أتدبر الأمر لا أجدني متأكدًا من هذا. هل تذكر كل تلك المعلومات المذهلة التي ذكرناها عن منع الجرائم قبل أن تحدث؟ في ذلك الصدد يمكن أن تقودنا الاحتمالات لأماكن مثيرة للفضول مقلقة في نفس الوقت. كيف يجدر بنا أن نتصرف عندما تُنبئنا تلك النماذج القائمة على الاحتمالات أن من يهرِّبون مخدر الميثامفيتامين من المكسيك هم على الأرجح رجال لاتينيون تتراوح أعمارهم بين الثامنة عشرة والثلاثين، ويقودون سيارات بيك أب بين التاسعة مساءً ومنتصف الليل، في حين أننا متأكدون من أن الغالبية العظمى من الرجال اللاتينيين الذين تنطبق عليهم تلك الأنماط لا يهرِّبون الميثامفيتامين؟ نعم، لقد استخدمتُ كلمة الأنماط لأنها تعتبر الوصف الأقل بريقًا للتحليلات التنبؤية التي وصفتُها بشكل برَّاق في الفصل السابق، أو هي على الأقل واحد من ملامحها المحتملة.

تخبرنا الاحتمالات بالأرجح والأقل أرجحية. نعم، فهي إحصائيات أولية، أو هي الأدوات التي وصفناها على مدى الفصول القليلة السابقة. لكنها أيضًا إحصائيات ذات دلالات اجتماعية. فإذا ما أردنا الإيقاع بالمجرمين العنيفين والإرهابيين ومهرِّبي المخدرات

* تم التعجيل بتطبيق تلك السياسة عام ٢٠١١ بعد أن قضت محكمة العدل بالاتحاد الأوروبي بأن التفاوت في تحصيل المبالغ بين الرجال والنساء يعد تفرقة على أساس الجنس.

وغيرهم ممن يحتمل أنهم سيسببون أضرارًا جسيمة، فيجب علينا أن نستخدم كل ما في جعبتنا من أدوات. ويمكن للاحتمالات أن تكون واحدة من تلك الأدوات. سيكون من السذاجة بمكان ألا نرى في الجنس والعمر والعرق والجنسية والدين وبلد المولد معًا ما يخبرنا بمعلومات تتعلق بإنفاذ القانون.

لكن ما نستطيع فعله أو يجدر بنا فعله بهذا النوع من المعلومات (على افتراض أنها ذات أهمية تنبؤية ما) هو مسألة فلسفية قانونية لا إحصائية. إننا نتلقى يوميًّا المزيد والمزيد من المعلومات عن المزيد والمزيد من الأشياء. هل لنا أن نفرق في المعاملة إذا أخبرَتنا البيانات بأننا حين التفرقة سنصيب أكثر مما نخطئ؟ (ذلك هو منشأ مصطلح «التفرقة الإحصائية» أو «التفرقة المنطقية»). فنفس التحليل، الذي يمكن أن يستخدم ليقرر أن أولئك الذين يشترون حبوب إطعام الطيور لن يتخلفوا عن أقساط بطاقات الائتمان على الأرجح (نعم هذا حقيقي)، يمكن أن يطبق على كل مناحي الحياة. لكن كم هو القدر المقبول من هذا التحليل؟ فإذا ما استطعنا وضع نموذج يتعرف على مهرِّبي المخدرات بدقة تبلغ ٨٠ من كل ١٠٠ حالة، فماذا يحدث للمساكين الذين يندرجون في قائمة العشرين بالمائة – لأن النموذج لن يتوقف عن التحرش بهم مرارًا وتكرارًا.

الفكرة الأعم هنا هي أن قدرتنا على تحليل البيانات قد تطورت بشكل يفوق ما نستطيع فعله بنتائج هذا التحليل. يمكنك أن تتفق أو تختلف مع قرار المفوضية الأوروبية لمنع التفرقة بين الجنسين في مبالغ التأمين، لكنني أعدك أنه لن يكون آخر قرار من نوعه. يروق لنا أن نرى الأرقام «حقائق جامدة باردة»، فإذا ما أجرينا الحسابات الصحيحة، فلا بد أن نصل إلى الإجابة الصحيحة. لكن المهم هنا والأكثر خطورة أننا أحيانًا نحسب الحسابات الصحيحة لكن ينتهي بنا الأمر أن نتخبط نحو اتجاهات خطرة. قد ندمر النظام المالي، أو قد نتحرش بشاب أبيض في الثانية والعشرين من عمره يقف في منعطف شارع معين في ساعة معينة من النهار، لأنه طبقًا للنموذج الإحصائي الذي نتبناه، من شبه المؤكد أنه قد وقف هنا ليبتاع المخدرات. رغم كل ما تتمتع به الاحتمالات من أناقة وتحديد، فإنه لا بد من التفكير في نوع الحسابات التي نحسبها والسبب من وراء حسابها.

الفصل السابع

أهمية البيانات

«تدخل قمامة، فتخرج قمامة»

في ربيع عام ٢٠١٢ نشر باحثون نتائج مذهلة في دورية ساينس المرموقة. خلص هذا البحث المذهل إلى أن ذكور ذباب الفاكهة إذا ما تعرضوا لصدٍّ متكرر من الإناث يلجئون إلى إغراق أحزانهم في الكحول. قامت صحيفة النيويورك تايمز بعرض تلك الدراسة في مقالة بصفحتها الأولى قائلة: «كانوا ذكورًا على أعتاب الشباب، وحاولوا مرات عديدة أن يجتذبوا مجموعة من الإناث الفاتنات الحائمات حولهم؛ لذا فقد فعلوا ما يفعله الكثير من الرجال عندما يُرفضون أكثر من مرة، لقد سكروا مستخدمين الكحول كمثبط لرغباتهم التي لم تشبع»[١]

إن هذا البحث قد طور من فهمنا لنظام المكافأة الموجود بالدماغ، وهو الأمر الذي من شأنه أن يساعدنا على إيجاد استراتيجيات جديدة للتعامل مع الاعتماد على المخدرات والكحوليات لدفن الأحزان. وصف الخبراء بإدمان المخدرات والكحوليات قراءة هذه الدراسة بأنها: «نظرة في التاريخ بهدف استكشاف أصول دائرة المكافأة التي تحت السلوكيات الأساسية كممارسة الجنس والأكل والنوم».

ولأنني لست خبيرًا في هذا المجال، فقد كان لي ردًّا فعل يختلفان قليلًا أحدهما عن الآخر تجاه ذكور ذباب الفاكهة المنبوذين أولئك. الأول أنهم قد أشعروني بالحنين إلى أيام الجامعة. والثاني أنهم حركوا الباحث الكامن في أعماقي وجعلوه يتساءل كيف لذباب الفاكهة أن يشرب حتى يسكر. أهناك حانة ذباب فاكهة مصغرة فيها أصناف من الأشربة المستخلصة من الفاكهة وبها ذكر ذباب فاكهة يعمل ساقيًا متعاطفًا؟ هل تصدح في خلفيتها موسيقى الريف الغربي؟ هل يحب ذباب الفاكهة موسيقى الريف الغربي من الأساس؟

لقد اتضح أن تصميم تلك التجربة كان بسيطًا بشكل شيطاني. فقد سمح لمجموعة معينة من ذكور ذباب الفاكهة بأن تتزاوج بحُرية مع الإناث العذارى. أما المجموعة الثانية

فقد أطلقت وسط مجموعة إناث قد سبق وتزاوجت بالفعل مع ذكور آخرين، ومن ثم لم تبالِ بعروضهم الغرامية. ثم قُدم لمجموعتي الذكور كليهما ماصات إطعام تمنحهم اختيارًا بين الطعام المعتاد لذباب الفاكهة والمكون من الخميرة والسكر، و«المشروبات القوية» المكونة من الخميرة والسكر وما نسبته ١٥ بالمائة من الكحول. حينها اتجه الذكور، الذين قضوا أيامًا يحاولون التزاوج مع الإناث غير المباليات، إلى اختيار المسكرات.

لكن المضحك في الأمر هنا هو أن لهذه النتائج دلالات لدى البشر. فهي تفترض وجود صلة بين التوتر والاستجابات الكيميائية في المخ والقابلية لتعاطي الكحول. لكن هذه النتائج لا تعتبر فتحًا إحصائيًّا، وإنما تعتبر فتحًا في مجال البيانات، والتي تجعل التحليل الإحصائي ممكنًا. لقد كان العبقري الذي أجرى هذه الدراسة يحاول إيجاد طريقة لخلق مجموعة من ذكور ذباب الفاكهة المشبعين جنسيًّا ومجموعة أخرى من المحبطين جنسيًّا، ثم يجد طريقة للمقارنة بين عاداتهم في شرب الكحول. وبعد أن حقق الباحث ذلك لم يكن حساب الأرقام أكثر تعقيدًا من مشروع معرض علوم لطالب في المدرسة الثانوية.

إن البيانات بالنسبة إلى الإحصاء بمثابة خط الهجوم القوي بالنسبة إلى لاعب خط وسط نجم. فدائمًا ما يكون أمام كل نجم في خط الوسط مجموعة ممتازة من المانعين. عادة لا ينال هؤلاء المانعون كثيرًا من الثناء الذي يستحقون. لكن مِن دونهم لن ترى لاعب خط وسط نجمًا أبدًا. تفترض معظم كتب الإحصاء أنك ستستخدم بيانات جيدة، تمامًا كما تفترض كتب الطبخ أنك لن تبتاع لحمًا نتنًا أو خضروات عطنة. لكن حتى أفضل وصفات الطبخ لا تقدر على أن تنقذ وجبة أعدت بمكونات فاسدة. وهكذا الحال مع الإحصاء، فلا يقدر أي تحليل أنيق على أن يصلح ما تفسده البيانات المعيبة. ومن هنا جاء تعبير «تدخل قمامة، فتخرج قمامة». إن البيانات تستحق الاحترام تمامًا كخط الهجوم.

إننا نطلب من بياناتنا ثلاثة أمور بشكل عام. أولًا نطلب عينة بيانات ممثلة لمجموعة كبيرة أو مجتمع كبير. إذا كنا نحاول أن نقيس آراء الناس تجاه مرشح سياسي بعينه، فسنحتاج أن نجري مقابلات مع عينة من الناخبين المحتملين الممثلين لجميع الناخبين في الدائرة الانتخابية المعنية. (وتذكر أننا لا نريد عينة ممثلة لجميع من يقطنون تلك الدائرة الانتخابية، وإنما نريدها ممثلة لمن يرجح أن يذهبوا للتصويت في الانتخابات). واحدة من أكثر النتائج قوة في مجال الإحصاء سوف نناقشها بعمق أكثر في الفصلين التاليين، ألا وهي أن الاستدلال الذي نقوم به بشكل صحيح من خلال عينات كبيرة نوعًا ما، على نفس القدر من الدقة التي نحصل عليها من خلال استقاء المعلومات من المجتمع بأكمله.

أسهل الوسائل التي نستطيع بها أن نجمع عينة ممثلة لمجتمع سكاني كبير هي انتقاء مجموعة فرعية عشوائية منه. (الغريب أن هذه العملية تعرف باسم عينة عشوائية بسيطة). وتعتبر الخاصية الأفضل في هذه المنهجية أن كل ملاحظة لها نفس الفرصة

نسبيًّا في أن تضمن في هذه العينة. أو للتعبير عن الأمر بشكل أكثر تحديدًا، نقول إنك إن خططت لأن تستقصي عينة عشوائية بسيطة مكونة من ١٠٠ شخص بالغ في حي يقطنه ٤٣٢٨ بالغًا، فلا بد أن تضمن منهجيتك أن يكون لكل عينة ممكنة مكونة من مائة شخص فرصةً متساوية لأن يتضمنها هذا الاستقصاء. ويمكن عمل هذا بأن تقوم، مثلًا، بانتقاء عشوائي لمائة رقم هاتف من دليل أرقام يحوي جميع أرقام من يعيشون في الحي، ثم مهاتفتهم. وهذا الانتقاء العشوائي (أو أي منهجية مشابهة) يقوم بدوره بضمان أن كل واحد من السكان الـ ٤٣٢٨ له نفس احتمالات أن يجد نفسه ضمن البالغين المائة الذين يتم استقصاؤهم. (دعنا نفترض افتراضًا غير واقعي أن كل هؤلاء السكان سوف يجيبون هواتفهم). دائمًا ما تشرح كتب الإحصاء هذه العملية من خلال التقاط كور زجاجية ملونة من داخل جَرَّة. (الواقع أن هذا قد يكون السياق الوحيد الذي تذكر فيه كلمة «جَرَّة» في إطار نظامي). فإذا تواجد في الجَرَّة الكبيرة ٦٠٠٠٠ كرة زجاجية زرقاء اللون، و٤٠٠٠٠ كرة زجاجية حمراء اللون، فإن أقرب التوليفات الممكنة لعينة المائة كرة زجاجية احتمالًا هي ٦٠ كرة زرقاء و٤٠ كرة حمراء. وإذا ما كررنا هذه العملية أكثر من مرة، فلا بد أن تحدث انحرافات تتباين من عينة لأخرى؛ فبعض العينات قد تحوي ٦٢ كرة زرقاء و٣٨ كرة حمراء، أو ٥٨ كرة زرقاء و٤٢ كرة حمراء. لكن احتمالات سحب عينة عشوائية تتباين كثيرًا عن نسب الألوان الموجودة في الجَرَّة قليلة جدًّا جدًّا.

لكن لنقرَّ الآن بوجود بعض الصعوبات العملية التي تتخلل هذا الأمر. فمعظم المجتمعات السكانية أكثر تعقيدًا من مجرد جَرَّة مليئة بالكرات الزجاجية. فكيف يمكن أن ينتقي أحدهم عينة عشوائية من الأمريكيين البالغين عن طريق مهاتفتهم تليفونيًّا؟ فحتى الحلول التي تبدو أنيقة، كالاتصالات الهاتفية العشوائية، قد تتخللها بعض العيوب. فبعض الأفراد (خاصة ذوي الدخول المنخفضة) قد لا يمتلكون هواتف، وبعضهم (خاصة ذوي الدخول العالية) قد يكونون ممن يفحصون الأرقام التي تتصل بهم قبل الإجابة فيختارون ألا يجيبوا. سوف يوجِز الفصل العاشر بعض الاستراتيجيات التي تستخدمها مؤسسات استطلاع الآراء للتغلب على هذه التحديات التي تواجهها عمليات انتقاء العينات بمختلف أنواعها (والتي ازدادت تعقيدًا مع التطور في استخدام الهواتف المحمولة). لكن الفكرة الأساسية هنا أن العينة المنتقاة بشكل صحيح سوف تأتي مشابة للمجتمع الذي انتقيت منه كاملة. كي تفهم الأمر جيدًا تخيل أخذ عينة من قدر مفعم بالحساء عن طريق غَرْف ملعقة واحدة منه. فإذا ما قلبت الحساء جيدًا فسوف تدلُّك تلك الملعقة على طعم الحساء كله الموجود في القدر.

إن أي كتاب في مجال الإحصاء سوف يتضمن تفاصيل أكثر بكثير تتعلق بمنهجيات انتقاء العينات. ومؤسسات استطلاع الرأي وشركات أبحاث السوق تنفق الوقت في دراسة

كيفية انتقاء بيانات من مجتمعات متعددة تكون سليمة وممثلة على الوجه الأمثل. لا بد لك هنا أن تضع في اعتبارك عددًا من الأمور المهمة: (١) إن العينة الممثلة أمر رائع مهم لأنها تفتح الباب أمام بعض أكثر الأدوات التي يقدمها الإحصاء فعالية. (٢) إن انتقاء عينة جيدة هو أمر أصعب مما يبدو عليه. (٣) الكثير من المسلَّمات الإحصائية الخطأ بشكل فاضح سببها تطبيق منهجيات إحصائية جيدة على عينات سيئة، لا العكس. (٤) الحجم مهم، وكلما كبرت صارت أفضل. سوف نسرد التفاصيل في الفصول اللاحقة، لكن لا بد أن ندرك أن العينات الكبيرة تساعد على التخلص السلس من أي تباينات شاذة. (فإذا ما غرفنا وعاءً من الحساء كان هذا أفضل من الاكتفاء بملء ملعقة). لكن علينا أن نحذر من أمر واحد وهو أن العينات الكبيرة لا تصحح أخطاء التوليف أو ما يعرف بـ «التحيزات». فالعينة السيئة سوف تظل دومًا عينة سيئة. فلا الحواسيب الفائقة ولا المعادلات المنمقة قادرة على أن تخلع على استطلاع الرأي الذي أجريته عن المرشحين الوطنيين للرئاسة المصداقيةَ إذا ما اعتمد هذا الاستطلاع فقط على مهاتفة المقيمين بواشنطن العاصمة. فقاطنو واشنطن العاصمة لا يصوتون كما يصوت باقي الأمريكيين، ومهاتفة ١٠٠٠٠٠ بدلًا من ١٠٠٠ لن تصلح المشكلة الأساسية التي تعيب ذلك الاستطلاع الذي أجريته. بل إن العينات الكبيرة المنحازة أسوأ، على حسب زعم بعضهم، من العينات الصغيرة المنحازة؛ لأنها تمنحنا إحساسًا زائفًا بالثقة في تلك النتائج.

الأمر الثاني الذي عادة ما ننشده من البيانات هو أن تمنحنا مصدرًا للمقارنة. أهذا الدواء الجديد أكثر فعالية من الدواء الحالي؟ هل مَن تلقوا تدريبًا مهنيًّا من المجرمين السابقين أقل عرضة للعودة إلى السجن من نظرائهم الذين لم يتلقوا مثل هذا التدريب؟ هل يُبلي تلاميذ المدارس المستقلة بلاءً أفضل من رفاقهم الذين يدرسون في المدارس العادية؟

هدفنا في هذه الحالات أن نجد مجموعتين من المدروسين يتشابهون بشكل كبير إلا في تطبيق أي «معالجة» نهتم بها. وفي سياق العلوم الاجتماعية يتسع معنى مفردة «معالجة» ليشمل كل شيء من الإحباط الجنسي الذي يشعر به ذكر ذبابة الفاكهة إلى ما يشعر به من يتعرض لخصومات ضريبية في دخله. وكما نفعل في أي تطبيق آخر للمنهج العلمي، نحاول أن نعزل تأثير تدخل محدد أو صفة محددة. كانت تلك عبقرية تجربة ذباب الفاكهة. ففيها توصل الباحثون إلى طريقة لخلق مجموعة ضابطة (الذكور الذين تزاوجوا) ومجموعة «معالجة» (الذكور الذين قوبلوا بالصد)، ومن ثم يكون الاختلاف الذي يحدث في سلوكياتهم في شرب الخمر معزوًّا إلى ما إذا كانوا قد أحبطوا جنسيًّا أم لا.

في العلوم الطبيعية والبيولوجية يتم خلق مجموعات المعالجة ومجموعات الضبط بشكل مباشر نسبيًّا. فيمكن للكيميائيين أن يُجروا هذه التغييرات من أنبوب اختبار لآخر، ثم يدرسوا اختلاف النتائج. ويمكن لعلماء الأحياء كذلك أن يفعلوا نفس الشيء مع أطباق

بيتري التي يستخدمونها. بل إن معظم الاختبارات التي تجرى على الحيوانات أكثر بساطة من حث ذباب الفاكهة على معاقرة الكحوليات. فيمكننا أن نجعل مجموعة من الفئران تتمرن على بساط جري، ثم نقارن ذكاءهم العقلي عند وضعهم في متاهة مع أداء مجموعة أخرى من الفئران لم تمارس مثل هذا التمرين. لكن عندما يشارك البشر في هذه العملية يزداد الأمر تعقيدًا. فالتحليل الإحصائي السليم يتطلب مجموعة ضابطة ومجموعة معالجة، لكننا لا نستطيع أن نجبر البشر على أن يفعلوا ما تفعله فئران المعامل. (وكثير من الناس لا يحبون حتى أن تجعل فئران المعامل تفعل تلك الأشياء). هل يسبب تكرار الإصابة بارتجاج مخي مشكلات في الجهاز العصبي في وقت لاحق من العمر؟ ذاك سؤال مهم جدًّا. فمستقبل رياضة كرة القدم الأمريكية (وربما رياضات أخرى) تتعلق بإجابته. لكنه ليس بسؤال يجاب عنه عن طريق تجارب تجرى على البشر؛ لذا فإلى أن نجد وسيلة نجعل بها ذبابات الفاكهة تعتمر خوذات وتقوم بدفاع منتشر، علينا أن نجد سبلًا أخرى لدراسة التأثير الطويل المدى للإصابات بالرأس.

واحد من أكثر التحديات تواترًا، والتي تواجه الأبحاث التي تتم على بشر، هو تكوين مجموعة معالجة ومجموعة ضابطة لا يختلفان إلا في كون واحدة من المجموعتين تتلقى معالجة لا تتلقاها الأخرى. ولذلك السبب صار «المعيار الذهبي» للبحث هو الانتقاء العشوائي، وهو عملية يوزع من خلالها البشر (أو المدارس أو المستشفيات أو أيًّا كان ما ندرسه) بشكل عشوائي، إما على مجموعة المعالجة وإما على المجموعة الضابطة. إننا لا نفترض أن جميع المدروسين متطابقون. وإنما نتخذ من الاحتمالات صديقًا (مرة أخرى)، فنفترض أن الانتقاء العشوائي سوف يقسم الخصائص ذات الصلة بين المجموعتين بتساوٍ – سواء كانت الخصائص التي يمكن ملاحظتها كالعرق ومستوى الدخل، أو الخصائص الضبابية التي لا نستطيع قياسها والتي لم نكن قد وضعناها في الاعتبار كالمثابرة والإيمان.

السبب الثالث الذي لأجله نجمع البيانات هو كما تصفه ابنتي المراهقة «لأيٍّ ما كان». فإننا أحيانًا لا نعرف ما سنصنع بتلك البيانات التي نجمعها، لكننا نجمعها لأننا نتوقع أن تصير ذات فائدة في مرحلةٍ ما. ويشبه ذلك ما يفعله محقق مسرح الجريمة عندما يطلب جمع كل الأدلة المادية الممكنة حتى يتم تصنيفها فيما بعد بغرض الاستدلال. بعض تلك الأدلة ستثبت فائدته وبعضها لا. ولو عرفنا منذ البداية أيها، بالضبط، ذو الفائدة لَما أجرينا التحقيق من الأساس.

إنك، على الأرجح، تعرف أن التدخين والسمنة هما عاملان مسببان لمرض القلب. لكنك، على الأرجح، لا تعرف أن دراسة طويلة المدى أجريت على قاطني بلدة فارمنجهام بولاية ماساتشوستس هي التي ساعدت على الربط بين هذين العاملين وذاك المرض. وفارمنجهام هي مدينة حضرية يقطنها نحو ٦٧٠٠٠ نسمة وتقع على بُعد حوالي عشرين ميلًا غرب

بوسطن. تعرف هذه المدينة بين غير الباحثين بأنها إحدى ضواحي بوسطن التي تتسم بعقارات ذات أسعار معقولة وتقع على مقربة من مجمع ناتيك التجاري الراقي. أما الباحثون فيعرفونها بأنها موطن دراسة فارمنجهام عن مرض القلب، وهي التي تعد واحدة من أنجح الدراسات الطولية وأكثرها تأثيرًا في تاريخ العلم الحديث.

والدراسات الطولية هي التي تجمع معلومات عن مجموعة كبيرة من الناس في مراحل زمنية كثيرة مختلفة، كأن تجمع كل عامين مثلًا. فنفس الخاضعين للتجربة يمكن أن يقابلوا بشكل دوري على مدى عشر أو عشرين أو خمسين عامًا، مما يخلق مخزونًا هائلًا من المعلومات. في دراسة فارمنجهام جمع الباحثون معلومات عن ٥٢٠٩ من البالغين الذين كانوا يقطنون فارمنجهام عام ١٩٤٨؛ بيانات تتعلق بالطول والوزن وضغط الدم، والخلفية التعليمية، والهيكل الأسري، والحمية الغذائية، والتدخين من عدمه وتعاطي المخدرات، وهلم جرًّا. لكن الأهم أن الباحثين ظلوا يجمعون بيانات متابعة من نفس الخاضعين للدراسة منذ المقابلة الأولى (وكذلك جمعوا بيانات عن أبنائهم لفحص العوامل الوراثية المتعلقة بمرض القلب). لقد استُخدمت بيانات فارمنجهام تلك في إصدار ما يقترب من ألفَي مقال أكاديمي منذ عام ١٩٥٠، من بينها ألف مقال بين عامي ٢٠٠٠ و٢٠٠٩.

لقد خرجت هذه الدراسات بنتائج كانت محورية في إدراكنا لمرض الشرايين القلبية، رغم أن كثيرًا من تلك النتائج صرنا نراه مسلَّمات حاليًّا؛ كأن تدخين السجائر، مثلًا، يزيد خطر الإصابة بمرض القلب (١٩٦٠)، أو أن النشاط البدني يقلل خطر الإصابة بمرض القلب والسمنة تزيده (١٩٦٧)، أو أن ارتفاع ضغط الدم يزيد من خطر الإصابة بسكتة دماغية (١٩٧٠)، أو أن ارتفاع مستوى البروتين الدهني المرتفع الكثافة (الذي يعرف باسم البروتين الجيد) يقلل من خطر الموت (١٩٨٨)، أو أن من يعاني آباؤهم أو إخوتهم من مرض الشرايين القلبية معرَّضون بشكل أكبر لأن يصابوا به (٢٠٠٤ و٢٠٠٥).

إن البيانات التي تجمعها الدراسات الطولية هي مكافئ سيارات الفيراري. تظهر قيمة تلك البيانات في مرحلة استكشاف العلاقات السببية التي قد تستغرق أعوامًا أو عقودًا لتتكشف. فمثلًا بدأت دراسة روضة أطفال بيري في ستينيات القرن الماضي على ١٢٣ طفلًا من الأمريكيين ذوي الأصول الأفريقية الذين تتراوح أعمارهم بين الثالثة والرابعة وينتمون إلى عائلات فقيرة. تم اختيار الأطفال الخاضعين للدراسة بشكل عشوائي؛ بحيث وُضع بعضهم ضمن مجموعة تلقت برنامجًا تعليميًّا مكثفًا لمرحلةِ ما قبل المدرسة، ووُضع الآخرون في مجموعة لم تتلقَّهُ. ثم شرع الباحثون يقيسون نتائج متعددة من المجموعتين على مدى العشرين عامًا التي تلت. جاءت النتائج تؤكد جدًّا على فوائد التعليم في الطفولة المبكرة. فأولئك التلاميذ الذين تلقوا ذلك البرنامج التعليمي المكثف حازوا مستوى ذكاء

أعلى في سن الخامسة. وكانوا الأقرب للتخرج في المدرسة الثانوية. وصارت دخولهم المادية أعلى في عمر الأربعين. وعلى النقيض منهم كان من لم يتلقوا ذلك البرنامج، فقد تعرض كثيرون منهم للاعتقال خمس مرات أو أكثر إبان بلوغهم الأربعين من عمرهم.

ليس من المستغرب أننا لا نستطيع جميعًا أن نمتلك سيارة فيراري، لذلك فإن المكافئ البحثي لسيارات التويوتا هي البيانات المستعرضة التي هي عبارة عن مجموعة من البيانات التي تجمع في مرحلة زمنية معينة. فمثلًا لو أراد علماء الأوبئة البحث في السبب وراء مرض جديد (أو تفشي مرض قديم)، فقد يلجئون لجمع بيانات عن المصابين؛ أملًا في إيجاد نمط يقود لمصدره. ماذا أكلوا؟ إلى أين سافروا؟ ماذا يجمعهم أيضًا؟ وقد يجمع الباحثون أيضًا بيانات عن الأشخاص الذين لم يصابوا بالمرض كي يوضحوا الفروق بين المجموعتين.

الواقع أن الحديث عن تلك البيانات الاستعراضية قد ذكرني بالأسبوع الذي سبق حفل زفافي وقت أن تحولت إلى جزء من مجموعة بيانات. كنت أعمل في كاتمندو بِنيبال عندما ظهرت نتيجة تحليلي إيجابية تؤكد إصابتي بمرض بالمعدة لا يفهمه الأطباء جيدًا يسمى «البكتيريا الزرقاء»، وهو مرض لم يكتشف إلا في مكانين فقط في العالم بأسره. تمكن الباحثون من عزل مسبب المرض لكنهم لم يتمكنوا من معرفة كُنه ذلك الكائن المجهري لأنه لم يعرف من قبل. عندما اتصلت ببلدي كي أُعْلم خطيبتي بمرضي أقررت بأنني أحمل أخبارًا سيئة. فلم تعرف الوسائل التي ينقل بها ذاك المرض ولم يعرف له علاج، ومن الممكن أن يسبب إعياءً عامًّا للمريض وأعراضًا جانبية أخرى تستمر من عدة أيام إلى عدة أشهر*. ولأننا لم يكن يفصلنا عن حفل العرس سوى أسبوع واحد، فقد كانت تلك مشكلة. هل سأسيطر تمام السيطرة على جهازي الهضمي بينما أخطو على ممر قاعة العرس؟ ربما.

لكنني حاولت أن أركز على الأخبار الجيدة. أولًا كان من المعروف أن تلك «البكتيريا الزرقاء» ليست مميتة. ثانيًا حظيت حالتي باهتمام خبراء في الأمراض الاستوائية، منهم من كان من بانجكوك البعيدة. كم هو أمر رائع! (كما أنني أحسنت صنعًا بالإصرار على تحويل دفة الحديث نحو خطط حفل الزفاف على غرار: «كفانا حديثًا عن مرضي الذي لا يرجى شفاؤه. زيديني حديثًا عن أزهار العرس»).

قضيت ساعاتي الأخيرة في كاتمندو أجيب على استبيان من ثلاثين صفحة يعج بأسئلة عن كل جانب من جوانب حياتي على غرار: أين أكلت؟ ماذا أكلت؟ كيف طبخت؟ هل مارست السباحة؟ أين أسبح عادة وكم مرة أذهب إلى المسبح؟ وكان كل من أصيبوا بالمرض يجيبون على ذات الأسئلة. ثم أخيرًا تم تحديد نوع مسبب المرض ذاك بأنه نوع

* في ذلك الوقت كان متوسط فترة استمرار المرض ثلاثة وأربعين يومًا مع انحراف معياري قدره أربعة وعشرون يومًا.

من أنواع البكتيريا الزرقاء التي يحملها الماء. (هذه البكتيريا زرقاء اللون لكنها النوع الوحيد من البكتيريا الذي يستمد طاقته من عملية التمثيل الضوئي كما في النباتات، لذلك أسماها البعض البكتيريا ذات اللون الأزرق المائل إلى الخضرة). وقد اتضح أن ذلك المرض يستجيب للعلاج بمجموعة من المضادات الحيوية التقليدية، لكن الغريب أنه لم يكن يتجاوب مع بعض الأنواع الحديثة منها. كل تلك الاكتشافات جاءت متأخرة فلم تفدني في شيء، لكن الحظ قد أسعدني وتعافيت بشكل أو بآخر. ويوم زفافي كنت أتحكم في جهازي الهضمي تحكمًا شبه كامل.

ووراء كل دراسة مهمة هناك بيانات جيدة تجعل التحليل ممكنًا. ووراء كل دراسة سيئة الأفضل أن تواصل القراءة. دائمًا ما يتكلم الناس عن «الكذب بالإحصائيات». إنني أزعم أن بعض أكثر الأخطاء الإحصائية فداحة تنطوي على كذب بالبيانات، حينها لا يشوب التحليل الإحصائي أي عوار، ولكن الحسابات تجرى على بيانات زائفة أو غير صحيحة. فيما يلي بعض أكثر الأمثلة شيوعًا على عبارة «تدخل قمامة، فتخرج قمامة».

الانحياز في الانتقاء. يقال إن بولين كيل الناقدة السينمائية المخضرمة بمجلة النيويوركر، قد صرحت عقب انتخاب ريتشارد نيكسون رئيسًا للولايات المتحدة: «لا يمكن أن يكون نيكسون قد ربح الانتخابات، فإنني لم أسمع عن أحد صوت له». هي مقولة مشكوك في صحتها لكنها مثال لطيف يُظهر كيف يؤدي سوء انتقاء العينات (حيث العينة هنا مؤلفة من مجموعة من الأصدقاء الليبراليين) إلى الركون إلى شريحة صغيرة مضللة من بين مجتمع أكثر ضخامة (من لهم حق التصويت في جميع أرجاء أمريكا). كما أنها تضع أمامنا السؤال الذي يتعين علينا طرحه على أنفسنا وهو: كيف انتقينا تلك العينة أو تلك العينات التي نشرع في تقييمها؟ فإذا لم يكن لكل فرد في هذا المجتمع ذي الصلة فرصة متساوية لأن يضمن داخل تلك العينة، فحينها ستواجهنا مشكلة في النتائج التي سنخرج بها من العينة أيًا كانت. هناك طقس يتعلق بالانتخابات الرئاسية يعقد في ولاية آيوا يسمى استطلاع آيوا الاقتراعي، وفيه يتجمع كل المرشحين الجمهويين المحتملين في آمس بولاية أيوا في شهر أغسطس قبل سنة من الانتخابات الرئاسية كي يخطبوا ود الناخبين، والذين يدفع كل منهم ثلاثين دولارًا كي يدلي بصوته في الاستطلاع. لكن استطلاع آيوا الاقتراعي لا ينبئنا بالكثير عن مستقبل المرشحين الجمهوريين. (فقد تنبأ الاستطلاع بثلاثة فقط من آخر خمسة مرشحين جمهوريين). لماذا؟ لأن أهل آيوا الذين يدفعون الثلاثين دولارًا كي يدلوا بأصواتهم في هذا الاستطلاع يختلفون عن بقية أهل آيوا من الجمهوريين، وأهل آيوا الجمهوريون يختلفون عن بقية المقترعين الجمهوريين في بقية أرجاء الدولة.

يمكن أن يحدث التحيز في الانتقاء بطرائق عدة. فأي استطلاع رأي يجرى على مستهلكين منتظرين في أحد المطارات سيكون منحازًا؛ لأن المسافرين جوًّا هم بشكل عام

أكثر ثراءً من غيرهم، أما الاستطلاع الذي يجرى في استراحة على الطريق السريع رقم ٩٠ ستشوبه ذات المشكلة ولكن بشكل عكسي. كلا الاستطلاعين منحازان؛ وذلك لأن من سيُقبلون على الإجابة عن أسئلة استطلاع الرأي في مكان عام يختلفون عن أولئك الذين لن يزعجوا أنفسهم بالإجابة. فإذا ما طرحت ١٠٠ سؤال على أناس جالسين في مكان عام في إطار استطلاع رأي قصير، فقَبِل ٦٠ شخصًا منهم أن يجيبوا عن الأسئلة، فإن هؤلاء الأشخاص الستين يختلفون اختلافًا بينًا، على الأرجح، عن الأربعين الذين انصرفوا عنك دون أن تلتقي أعينهم بعينيك.

من أشهر الأخطاء الإحصائية الفادحة على مر التاريخ كان استطلاع الرأي الفضيحة الذي أجرته مجلة ليتراري دايجست عام ١٩٣٦، والذي كانت وراءه عينة منحازة. في ذلك العام كان حاكم ولاية كانزاس آلف لاندون الجمهوري الانتماء الحزبي يخوض انتخابات رئاسة الولايات المتحدة الأمريكية أمام فرانكلين روزفلت الديمقراطي، الذي كان يحتل وقتها كرسي الرئاسة. حينها قامت مجلة الليتراري دايجست، تلك المجلة الإخبارية التي كانت تصدر أسبوعيًّا في ذلك الوقت، بإجراء استطلاع رأي أرسلته عبر البريد لمشتركيها ولمالكي السيارات والهواتف الذين يمكن التحصل على أرقام هواتفهم من السجلات العامة. أخبر الجميع أن استطلاع الرأي ذلك قد ضم عشرة ملايين مصوت محتمل، وهو ما يعد عينة ضخمة ذات رقم فلكي. كلما زادت ضخامة استطلاعات الرأي ذات العينات الجيدة، زادت جودتها؛ وذلك لأن هامش الخطأ يتقلص. وكلما زادت ضخامة استطلاعات الرأي ذات العينات السيئة، زاد عدد القمامة الداخلة فيها وزادت رائحتها نتانة. توقعت المجلة فوز لاندون على روزفلت بأصوات الجماهير بفارق تبلغ نسبته ٥٧ بالمائة. لكن ما حدث هو أن روزفلت قد فاز بأغلبية ساحقة بلغت ٦٠ بالمائة من أصوات الناخبين وست وأربعين ولاية من ثمانية وأربعين ولاية في المجمع الانتخابي. لقد كانت عينة مجلة ليتراري دايجست «قمامة داخلة»: فقد كان المشتركون بالمجلة أكثر ثراء من الأمريكي العادي، ومن ثم فإنهم أميل لأن يصوتوا للجمهوريين، وكذلك كان حال العائلات التي تمتلك هواتف وسيارات عام ١٩٣٦ [٢].

إننا كذلك معرضون لأن نقع في نفس المشكلة الأساسية عندما نقارن ناتجَي مجموعة المعالجة بمجموعة الضبط إذا كانت آلية تصنيف الأفراد في إحدى المجموعات عشوائية. دعنا نتأمل تلك النتيجة الطبية التي تم التوصل لها حديثًا عن الأعراض الجانبية لعلاج سرطان البروستاتا. هناك ثلاث علاجات شائعة لسرطان البروستاتا؛ وهي الإزالة الجراحية للبروستاتا، والعلاج الإشعاعي، والمعالجة القريبة (والتي تتم عن طريق زرع «بذور» مشعة قرب الورم السرطاني)[٣]. من الأعراض الجانبية الشائعة لعلاج سرطان البروستاتا هو العجز الجنسي؛ لذلك صار الباحثون يوثقون الوظائف الجنسية للرجال الذين تلقوا

أيًّا من العلاجات الثلاثة. فبعد دراسة أجريت على ١٠٠٠ رجل، خلص الباحثون إلى أنه بعد عامين من العلاج استعاد ٣٥ بالمائة من الرجال الذين كانوا في المجموعة التي خضعت للعلاج الجراحي قدرتهم على المعاشرة الجنسية، مقارنة بـ ٣٧ بالمائة من مجموعة العلاج الإشعاعي و٤٣ بالمائة من مجموعة المعالجة القريبة.

هل يمكننا أن ننظر في تلك البيانات فنفترض أن المعالجة القريبة هي العلاج ذو الاحتمالات الأقل في تدمير قدرة الرجال الجنسية؟ لا، لا، لا. فقد قام واضعوا هذه الدراسة بوضع تحذير صريح من أننا لا نستطيع أن نخلص إلى أن المعالجة القريبة هي الأقدر على الحفاظ على الوظائف الجنسية للرجال؛ وذلك لأن أولئك الرجال الذين تلقوا هذا العلاج هم بشكل عام أصغر عمرًا وأعلى لياقة من الذين تلقوا العلاجات الأخرى. لقد كان الهدف من هذه الدراسة هو مجرد توثيق مستويات الأعراض الجانبية من ناحية الوظائف الجنسية التي يسببها كل نوع من أنواع العلاجات الثلاثة.

وهناك نوع ذو صلة من أنواع الانحياز يعرف باسم الاختيار الذاتي للانحياز، وهو الذي ينشأ عندما يتطوع أفراد لأن يضعوا أنفسهم في مجموعة المعالجة. فمثلًا يختلف السجناء الذين يتطوعون في مجموعات علاج إدمان المخدرات عن بقية السجناء؛ وذلك لأنهم تطوعوا لأن يكونوا في برنامج علاج إدمان المخدرات. إذا كان المشاركون في هذا البرنامج أكثر قابلية لأن يتجنبوا العودة إلى السجن عقب إطلاق سراحهم من بقية السجناء، فهو أمر عظيم – لكن هذا لا يخبرنا بأي شيء عن فائدة برنامج علاج الإدمان ذلك. فقد يكون هؤلاء المساجين السابقون قد غيروا من نمط حياتهم لأن البرنامج قد ساعدهم على أن يقلعوا عن المخدرات. أو قد يكونون غيروا من نمط حياتهم بسبب عوامل أخرى تضافرت هي أيضًا فجعلتهم أكثر قابلية لأن يتطوعوا في برنامج علاج إدمان المخدرات (كأن تكون لدى السجين رغبة قوية جدًّا في ألا يعود إلى السجن مرة أخرى). لا نستطيع أن نفصل التأثير العارض لأحد العوامل (برنامج علاج الإدمان) عن العامل الآخر (كون المرء شخصًا طيبًا بشكل يجعله يتطوع في برنامج علاج الإدمان).

الانحياز المطبوع. إن النتائج الإيجابية هي الأكثر احتمالات لأن تطبع وتنشر عن النتائج السلبية، وهو ما قد يزيف الخلاصات التي نراها. لنفترض أنك قد أجريت لتوِّك دراسة طولية دقيقة استخلصت منها نتيجة حاسمة مفادها أن ممارسة ألعاب الفيديو لا تَحُول دون الإصابة بسرطان القولون. لقد قمت بتتبع ١٠٠٠٠٠ أمريكي على مدى عشرين عامًا، هؤلاء المشاركون في التجربة الذين ينفقون الساعات في ممارسة ألعاب الفيديو لهم نفس المعدل تقريبًا للإصابة بسرطان القولون الذي يتسم به من لا يمارسونها على الإطلاق. لنفترض أن منهجيتك لا تشوبها شائبة. ما تلك الدورية الطبية المرموقة التي ستوافق على نشر نتائجك تلك؟

لن تجد، وذلك لسببين. أولهما أنه لا يوجد دليل علمي قوي يدفع للاعتقاد بأن لعب ألعاب الفيديو له أي تأثير على سرطان القولون؛ لذا فسبب إجرائك لهذه الدراسة غير واضح. وثانيهما والأكثر أهمية هو أن حقيقة كون شيءٍ ما لا يمنع حدوث السرطان لا يعد بالنتيجة المثيرة للاهتمام. فالنتائج السلبية ليست نتائج جذابة في الطب أو في غيره.

التأثير النهائي هنا هو أن البحث الذي نراه أو لا نراه يشوبه الزيف. لنفترض أن إحدى زميلاتك في الدراسات العليا قامت بإجراء دراسة طولية تفاضلية، فوجدت أن من ينفقون الساعات في ممارسة ألعاب الفيديو تنخفض نسبة إصابتهم بسرطان القولون. هنا يصير الأمر مثيرًا للاهتمام. فهذا بالضبط هو نوع النتائج الذي يجذب اهتمام الدوريات الطبية والصحافة الشعبية والمدونين وصانعي ألعاب الفيديو (الذين سيضعون ملصقات على منتجاتهم تشرح الفوائد الطبية لهذه الألعاب). ولن يمضي وقت طويل حتى نجد أولئك الأمهات المغاليات في حماية أبنائهن في جميع أرجاء الدولة، «يدفعن» السرطان عنهم بأن ينزعن الكتب من أيديهم ويجبرنهم على الانغماس في ألعاب الفيديو.

واحدة من أهم الفِكَر المتواترة في الإحصاء هي، ولا شك، أن الأمور غير المعتادة تحدث بين فينة وأخرى على سبيل الصدفة. إنك إذا أجريت مائة دراسة، فإن واحدة منها قابلة لأن تخرج بنتائج ما هي إلا محض هراء، كالارتباط الإحصائي بين ممارسة ألعاب الفيديو والإصابة بسرطان القولون. تتمثل المشكلة فيما يلي: لن تُنشر الدراسات التسع والتسعون التي لم تجد أي رابط بين ألعاب الفيديو وسرطان القولون؛ وذلك لأنها غير مثيرة للاهتمام لهذه الدرجة. أما الدراسة الوحيدة التي ستنجح في إيجاد ذلك الرابط الإحصائي، فسوف تُطبع وتُنشر وتَجذب اهتمامًا هائلًا. إن الانحياز لا ينبع من الدراسات نفسها، وإنما من المعلومات الزائفة التي تصل إلى العامة. فمن يقرأ الكتب العلمية التي تتناول ألعاب الفيديو ومرض السرطان، لن يجد إلا دراسة وحيدة تزعم أن ألعاب الفيديو تقي من السرطان. لكن الواقع أن تسعًا وتسعين دراسة من أصل مائة لم تجد بين الظاهرتين أي رابط.

نعم، ذلك مثال عبثي – لكن المشكلة حقيقية عويصة. افتتحت صحيفة النيويورك تايمز مقالًا عن الانحياز المطبوع الذي يحيط العقاقير التي تعالج الاكتئاب بهذه الجملة: «إن مُنتِجِي مضادات الاكتئاب، مثل بروزاك وباكسيل، لم ينشروا نتائج حوالي ثلث تجارب العقاقير التي أجروها ليحظوا بموافقة الحكومة، مضللين بذلك الأطباء والمستهلكين عن الفعالية الحقيقية للعقَّار»[٤]. لقد اتضح أن ما نسبته ٩٤ بالمائة من الدراسات التي خرجت بنتائج إيجابية عن فعالية هذه العقاقير هي التي نشرت، في حين لم تنشر من الدراسات التي خرجت بنتائج غير إيجابية إلا ما نسبته ١٤ بالمائة فقط. بالنسبة إلى المرضى الذين يعانون الاكتئاب يعتبر هذا أمرًا مهمًّا. لكن إذا ما وضعت جميع الدراسات في الاعتبار فسنجد مضادات الاكتئاب أفضل من العقاقير الوهمية التي تستخدم في التجارب «بهامش بسيط».

للتغلب على تلك المشكلة صارت الدوريات الطبية تشترط أن تكون أي دراسة تجرى في بداية المشروع قابلة للنشر في وقت لاحق. يساعد هذا محرري الدورية الطبية على أن يستدلوا على نسبة النتائج الإيجابية إلى النتائج غير الإيجابية. فإذا ما تم تسجيل مائة دراسة تخطط لاختبار أثر ممارسة التزلج على مرض القلب، ثم لم تقدم منها للنشر سوى واحدة ذات نتائج إيجابية، حينها يستطيع المحررون أن يستنتجوا وجود دراسات أخرى خرجت بنتائج غير إيجابية (أو على الأقل يحققوا في هذه الإمكانية).

انحياز الذاكرة. الذاكرة شيء رائع رغم أنها لا تعتبر مصدرًا عظيمًا للبيانات. إننا نتسم بدافع إنساني طبيعي لإدراك الحاضر على أنه نتيجة منطقية لما حدث في الماضي – سببًا وأثرًا. تكمن المشكلة في أن ذاكرتنا تصير «هشة بشكل منهجي» عندما نحاول أن نفسر بعض النواتج السيئة أو الجيدة التي تحدث في الوقت الحالي. لنتأمل تلك الدراسة التي بحثت في العلاقة بين الحمية الغذائية والإصابة بمرض السرطان. في عام ١٩٩٣ قام أحد باحثي جامعة هارفارد بجمع بيانات عن مجموعة من النساء اللواتي أصبن بسرطان الثدي، ومجموعة أخرى من النساء في ذات العمر لا يعانين هذا المرض. تم سؤال النساء من المجموعتين عن الحمية الغذائية التي كنَّ يتبعْنها في مراحل مبكرة من حياتهن. خرجت الدراسة بنتائج واضحة أفادت بأن النسوة اللاتي أُصبْنَ بسرطان الثدي كنَّ أميل لاتباع حمية غذائية ذات دهون عالية عندما كن أصغر سنًّا.

لكن تلك الدراسة لم تكن تتناول كيفية تأثير الحمية الغذائية على احتمالات الإصابة بالسرطان. وإنما كانت تتناول الكيفية التي تؤثر بها الإصابة بالسرطان على تذكر المرأة لحميتها الغذائية التي اتبعتها في بداية حياتها. فجميع النسوة اللاتي خضعن للدراسة كن قد أجبن عن أسئلة استقصاء عن حميتهن الغذائية قبل أن تصاب أيٌّ منهن بالسرطان. لكن النتيجة المدهشة كانت أن النساء اللاتي أصبن بسرطان الثدي صورت لهن ذاكرتهن أن حميتهن كانت تحتوي دهونًا عالية، لكنها في الواقع لم تكن بهذا العلو، لكن هذا لم يحدث مع اللاتي لم يصبن بالسرطان. وصفت صحيفة النيويورك تايمز تلك الطبيعة الخبيثة التي يتسم بها انحياز الذاكرة بقولها:

> عندما يشخَّص مرض امرأة بأنه سرطان الثدي فهذا الأمر لا يغير فقط حاضرها ومستقبلها، وإنما يغير ماضيها كذلك. لقد قررت النساء اللاتي أصبن بسرطان الثدي (بشكل لا واعٍ) أن الحمية الغذائية العالية الدهون مسئولة على الأرجح عن مرضهن، وأنهن يذكرن (بشكل لا واعٍ) أنهن كن يستهلكن الكثير من الدهون. لقد كان نمطًا مألوفًا بشكل مؤلم لكل من يعرف تاريخ مرضه السيء السمعة، فهؤلاء النسوة والآلاف من قبلهن قد فتشن في ذاكرتهن عن سبب لمرضهن، ومن ثم استدعين ذلك السبب إلى ذاكرتهن[٥]

وانحياز الذاكرة هو واحد من الأسباب التي تمنح الأفضلية للدراسات الطولية على دراسات البيانات الاستعراضية. ففي الدراسات الطولية تُجمع البيانات بشكل معاصر. فيمكن سؤال المشاركين في التجربة وقت أن يكونوا في عمر الخامسة عن آرائهم في المدارس. ثم يمكن أن نعيد زيارة المدرسة وقت أن يبلغوا الثالثة عشرة لنرى مَن منهم ترك الدراسة. أما في دراسة البيانات الاستعراضية التي تُجمع فيها البيانات في مرحلة زمنية معينة، فيتعين علينا أن نسأل تلميذًا في المدرسة الثانوية يبلغ من العمر ثمانية عشر ربيعًا كيف كان يرى المدرسة وقت أن كان في الخامسة من عمره، وهو الأمر الذي لا يمنحنا نفس الثقة في البيانات التي تمنحنا إياها الدراسة الأخرى.

الانحياز للبقاء. لنفترض أن مدير مدرسةٍ ما قد أعلن أن نتائج اختبارات مجموعة معينة من التلاميذ قد تحسنت بشكل مضطرد على مدى أربع سنين. فقد جاءت نتائجهم في السنة الثانية أفضل منها في سنتهم الأولى. وجاءت نتائج السنة الثالثة أفضل من سابقتها، وجاءت نتائج السنة الأخيرة أفضلها قاطبة. سنفترض عدم حدوث أي غش في الاختبارات، وحتى أن أحدًا لم يقم باستخدام الإحصائيات الوصفية أي استخدام مبتكر، فنقول إن هذا الرهط من التلاميذ صار يُبلي في كل سنة بلاءً أفضل مما فعل في السنة التي سبقتها بأي مقياس كان، كالمتوسط والوسيط والنسبة المئوية لكل صف من الصفوف الدراسية، وهلم جرًّا.

ماذا عساك تفعل حيال ذلك؟ هل ستخلع على هذا المدير لقب «أفضل مدير هذا العام»، أم إنك ستطلب المزيد من البيانات؟

ما سأفعله أنا هو أنني سأطلب المزيد من البيانات، فإنني أشتمُّ رائحة انحياز للبقاء، وهو الأمر الذي يحدث عندما تخرج بعض المشاهدات أو الكثير منها عن العينة مغيِّرةً تركيب باقي المشاهدات، ومن ثم مغيِّرةً نتائج أي تحليل. لو افترضنا أن هذا المدير سيء للغاية ولا يتعلم التلاميذ في مدرسته أي شيء وفي كل عام يتسرب نصفهم من التعليم، فسوف يجعل هذا نتائج اختبارات التلاميذ الباقين في المدرسة تبدو ممتازة دون أن يتحسن مستوى أيٍّ منهم. لو وضعنا افتراضًا منطقيًّا ينص على أن أسوأ التلاميذ (ذوي أسوأ النتائج في الاختبارات) هم الأكثر قابلية لأن يتسربوا من التعليم، فحينها سوف يَأْخُذ متوسطُ نتائج اختبارات التلاميذ الباقين في الارتفاع بشكل مطرد، خاصة مع استمرار تسرب تلاميذ آخرين من التعليم. (إذا كنت في غرفة مليئة بأناس متبايني الأطوال، ثم طلبت من قصار القامة أن يغادروا الغرفة، فإن هذا سيرفع من معدل الطول لدى من في الغرفة، لكنه لن يجعل أيًّا منهم يزداد طولًا).

لقد تبنت صناعة صناديق الاستثمار المشترك بشكل عنيف (وغادِر) ذلك الانحياز للبقاء كي تجعل عوائدها تبدو أفضل في عيون المستثمرين لديها. دائمًا تقيس صناديق

الاستثمار المشترك أداءها طبقًا لمؤشر أساسي لأداء السندات، ألا وهو مؤشر ستاندرد آند بوورز ٥٠٠، والذي هو قائمة للشركات الخمسمائة الكبرى في أمريكا*. فإذا وصل مؤشر ستاندرد آند بوورز ٥٠٠ إلى ٥٫٣ بالمائة لهذا العام، فحينها سيتصدر أي صندوق استثمار مشترك يتعدى ذلك الرقم القائمة، وسيتذيلها أي آخر ينخفض عنه. واحدة من الطرق السهلة الزهيدة الكلفة التي يستطيع أن يتخذها أي مستثمر لا يرغب في أن يدفع نقودًا لأي مدير لصندوق استثمار مشترك هي أن يشتري صندوق استثمار مشترك من مؤشر قائمة ستاندرد آند بوورز ٥٠٠، وهو صندوق استثمار مشترك يقوم بكل بساطة بابتياع أسهم في كل الشركات الخمسمائة المضمنة في القائمة. يحب مديرو صناديق الاستثمار المشترك أن يظنوا أنفسهم مستثمرين ذوي دهاء قادرين على استخدام معارفهم لانتقاء السندات التي سوف تبلي بلاءً حسنًا يفوق نظيره لدى صندوق استثمار قائمة الخمسمائة البسيط. لكن الواقع أنه من الصعب نسبيًّا مواصلة التفوق على صندوق قائمة الخمسمائة لفترة زمنية كبيرة. (قائمة ستاندرد آند بوورز ٥٠٠ هو في جوهره متوسط لجميع السندات الضخمة المتداولة، لذلك فإننا من خلال الحسابات الرياضية لنا أن نتوقع أن نحو نصف صناديق الاستثمار المشترك النشطة سوف يتفوق على مؤشر ستاندرد آند بوورز ٥٠٠ في أي عام من الأعوام، بينما سوف يتأخر نصفهم عنه). بالطبع ليس هذا بالموقف الجيد بالنسبة إلى مديري صناديق الاستثمار المشترك أن ينهزموا أمام قائمة عجماء لا تفعل سوى أنها تشتري ٥٠٠ سهم ثم تتمسك بها. لا تحليلات، لا تنبؤات منمقة، وأخيرًا لا مصاريف إدارية باهظة، وهو ما سيثلج صدور المستثمرين.

ماذا تفعل شركات صناديق الاستثمار المشترك؟ إنها تجد الخلاص في البيانات الزائفة. سأريك كيف تسود تلك الشركات السوق دون أن «تسود السوق» حقًّا. تفتح أي شركة كبيرة لصناديق الاستثمار المشتركة الكثير من تلك الصناديق النشطة (بمعنى أن الخبراء يقومون بانتقاء السندات باستراتيجية معينة). لأجل هذا المثال دعنا نفترض أن تلك الشركة قد فتحت عشرين صندوقًا جديدًا لكلٍّ منها فرصة ٥٠ بالمائة لأن يتفوق على صندوق ستاندرد آند بوورز ٥٠٠ في أي عام من الأعوام. (يتوافق هذا الافتراض مع بيانات المدى الطويل). هنا تنص الاحتمالات الأساسية على أن عشرة فقط من تلك الصناديق الجديدة سوف تتفوق على شركات ستاندرد آند بوورز الخمسمائة خلال العام الأول، وسوف تكرر خمسة منها هذا في العامين التاليين، وسوف يفعل اثنان أو ثلاثة نفس الأمر خلال ثلاثة أعوام تالية.

* مؤشر ستاندرد آند بوورز ٥٠٠ هو مثال ممتاز لما ينبغي للقوائم أن تفعل. وهذه القائمة مؤلفة من أسعار أسهم الشركات الخمسمائة الكبرى في الولايات المتحدة، كلٌّ منها ذات ثقل يتوافق مع قيمتها السوقية (وهكذا فإن الشركات الأكبر تحظى بثقل أكبر من الشركات التي تصغرها في هذه القائمة). تعتبر هذه القائمة مقياسًا بسيطًا دقيقًا لما هو حادث لأسعار أسهم أضخم الشركات الأمريكية في أي وقت كان.

هنا يأتي وقت الجزء الحاذق. عند هذه النقطة تكون صناديق الاستثمار المشترك الجديدة ذات العوائد غير الجيدة نسبيًّا قد أغلقت عن صمت (وأدمجت أصولها في صناديق أخرى لم تغلق). عندها تستطيع الشركة أن تقوم بحملة ترويجية مكثفة تعلن أن صندوقين أو ثلاثة قد «استمروا متفوقين على صندوق ستاندرد آند بوورز ٥٠٠» – حتى لو كان تفوقه ذلك هو مجرد مكافئ لإصابة وجه الرأس ثلاث مرات متتالية من رمي العملات. والأداء المتوقع من هذه الصناديق هو أن ترتد إلى المتوسط، لكن بعد أن يتكدس المستثمرون فيها. وعدد شيوخ الاستثمار في صناديق الاستثمار المشتركة تلك، والذين استمروا في التفوق على صندوق ستاندرد آند بوورز ٥٠٠ على مدًى زمني طويل، هو عدد صغير للغاية*.

انحياز المستخدم الصحي. إن من يتعاطون الفيتامينات بشكل منتظم هم الأكثر قابلية لأن يكونوا أصحاء – وذلك لأنهم من نوع البشر الذين يتعاطون الفيتامينات بشكل منتظم. وكون الفيتامينات لها نصيب من صيرورة أولئك البشر على هذه الحالة هو مسألة منفصلة. لنتأمل التجربة الفكرية التالية. لنفترض أن مسئولي الصحة العامة قد أعلنوا عن نظرية مفادها أن على جميع الآباء الجدد أن يرتدوا ثياب نوم قرمزية اللون حين يضعون أبناءهم للنوم؛ لأن هذا يساعد على تحفيز نمو أدمغتهم. ثم بعد مضي عشرين عامًا أكدت دراسة طولية إلى أن ارتداء ثياب النوم القرمزية في الطفولة هو أمر ذو ارتباط إيجابي كبير بالنجاح في الحياة. فنجد مثلًا أن ٩٨ بالمائة ممن التحقوا بجامعة هارفارد كانوا يرتدون ثياب النوم القرمزية في طفولتهم (وكثير منهم ما زالوا يرتدونها حتى الآن) مقارنة بثلاثة في المائة منهم فقط الذين صاروا نزلاء بسجن ولاية ماساتشوستس.

بالطبع ليس ارتداء ثياب النوم القرمزية بالأمر المهم، لكن ارتداء الآباء هذه الثياب هو ما يهم. وحتى حينما نحاول أن نثبت عوامل أخرى كمستوى تعليم الأبوين، سنظل أمام اختلافات غير ملحوظة بين أي أبوين مهووسين بارتداء ثياب النوم القرمزية وغيرهما. يفسر جاري توبيس، الكاتب في مجال الصحة بصحيفة النيويورك تايمز، الأمر بقوله: «إن المشكلة تتمثل بكل بساطة في كون أولئك الذين ينخرطون بالتزام في الأنشطة التي تخدمهم – كتعاطي العقاقير الموصوفة لهم أو تناول ما يرونه صحيًّا من الأطعمة – يختلفون بشكل جوهري عمَّن لا يفعلون»[٦] هذا التأثير من الممكن أن يربك أي دراسة تحاول أن تقيم التأثير الحقيقي للأنشطة التي يعتقد أنها صحية، كممارسة الرياضة بانتظام أو تناول الكرنب. نعتقد أننا نقارن التأثيرات الصحية لحميتين غذائيتين: تناول الكرنب في مقابل عدم تناوله. لكن الواقع أننا إذا لم نكن قد وزعنا أشخاصًا معينين على

* لقراءة نقاش ثري عن السبب الذي لأجله ينبغي لك أن تشتري صناديق تلك القائمة ولا تحاول تصدر السوق طالع كتاب A Random Walk Down Wall Street لبرتون مالكيل أستاذي السابق.

مجموعتي المعالجة والضبط بشكل عشوائي، فإننا لا نفعل سوى أن نقارن بين نوعي حمية يتبعهما صنفان مختلفان من البشر. فما أمامنا الآن هو مجموعة معالجة تختلف عن مجموعة الضبط في جانبين لا واحد فقط.

إذا كان الإحصاء من قبيل أعمال التحريات، فإن البيانات هي الأدلة. قضت زوجتي عامًا معلمة في مدرسة ثانوية بولاية نيوهامبشاير الرعوية. وهناك تم إلقاء القبض على أحد تلامذتها لأنه اقتحم مخزن المعدات وسرق منه بعضها. استطاعت الشرطة أن تحل لغز السرقة لسببين: (١) أن السماء كانت قد أمطرت ثلجًا لتوِّها، فاستطاعوا تتبع آثار إطار السيارة الممتد من المستودع إلى منزل التلميذ. (٢) أن الأدوات المسروقة قد وجدت داخل المنزل. الأدلة الجيدة تساعد كثيرًا.

عليك أن تحب البيانات الجيدة، لكن عليك أولًا أن تتحصل على تلك البيانات الجيدة، وهو أمر أكثر صعوبة مما تظن.

الفصل الثامن

مبرهنة النهاية المركزية

ليبرون جيمس* الإحصاء

أحيانًا ما يبدو الإحصاء كالسحر، فبه نستطيع استخلاص نتائج قوية كاسحة من بيانات قليلة نسبيًّا. إذ قد نتمكن من التوصل لحقائق ذات معنًى متعلقة بالانتخابات الرئاسية الأمريكية من مجرد مهاتفة ألف من الناخبين. ويمكننا أن نفحص مائة قطعة من قطع صدور الدجاج مفتشين عن السالمونيلا في أحد مصانع تعبئة الدجاج، ومن تلك العينة فقط نخلص إلى أن المصنع بأكمله آمن أو ليس كذلك. من أين أتت تلك القوة التعميمية الخارقة؟

يأتي القدر الأكبر منها من مبرهنة النهاية المركزية، والتي هي للإحصاء كليبرون جيمس لكرة السلة – هذا إذا كان ليبرون جيمس يعمل أيضًا عارض أزياء وأستاذًا بجامعة هارفارد وكان قد حاز جائزة نوبل كذلك. ومبرهنة النهاية المركزية هي «مصدر القوة» للعديد من الأنشطة الإحصائية التي يتخللها استخدام العينات في الوصول لاستدلالات عن المجتمعات الكبيرة (كاستطلاعات الرأي واختبارات السالمونيلا). وأنواع الاستدلالات تلك التي قد تبدو ساحرة هي في الواقع لا تتعدى جمعًا بين أداتين قد استكشفناهما من قبل، ألا وهما الاحتمالات والانتقاء الجيد للعينات. وقبل أن نخوض في آليات عمل مبرهنة النهاية المركزية (والتي هي ليست عسيرة لهذه الدرجة)، إليك هذا المثال الذي سيجعل الأمر يبدو لك منطقيًّا.

لنفترض أنك تعيش في مدينة تستضيف ماراثونًا. وفيه سيتنافس عداءون من جميع أرجاء العالم، وهو ما يعني أن الكثير منهم لا يتكلمون الإنجليزية. يقتضي التنظيم اللوجيستي من العدَّائين أن يسجلوا أسماءهم صباح يوم السباق، ثم يوزعوا بشكل

* أحد أشهر لاعبي دوري كرة السلة للمحترفين، ويلعب ضمن صفوف نادي لوس آنجيليس وكذلك هو واحد من نجوم منتخب الولايات المتحدة لكرة السلة (المترجم).

عشوائي على حافلات لتأخذهم إلى خط بداية السباق. لكن لسوء الحظ حدث أن تاهت واحدة من تلك الحافلات بينما كانت متجهة إلى السباق. (لأجل هذا المثال، عليك أن تفترض أن لا أحد في الحافلة يحمل هاتفًا محمولًا، وأن السائق ليست لديه آلة تحديد الاتجاه GPS، افعل هذا وإلا سيتعين عليك أن تُجري الكثير من الحسابات المرهقة). ولأنك من القادة المدنيين، فقد انضممتَ إلى فريق البحث عنها.

ولأنك رجل محظوظ، فقد عثرت في حافلة معطلة قرب منزلك بها مجموعة كبيرة من الركاب الأجانب المستائين وليس من بينهم من يتحدث الإنجليزية. لا بد أنها الحافلة التائهة، سوف تصير بطلًا، لولا أمر واحد يبعث فيك الشك، ألا وهو أن ركاب هذه الحافلة، في الواقع، ضخام جدًّا. فبنظرة سريعة قدَّرت أن متوسط الوزن في مجموعة الركاب تلك يبلغ ٢٢٠ رطلًا. ولا يمكن أن تكون مجموعة عشوائية من عدَّائي الماراثون بهذا الوزن الثقيل. ومن ثم تخاطب مقر الماراثون عبر جهاز اللاسلكي قائلًا: «لا أظنها الحافلة المنشودة، واصلوا البحث».

المزيد من التحليل سوف يؤكد انطباعك الأولي ذلك. فعندما يصلك المترجم ستكتشف أن تلك الحافلة المعطوبة كانت تتجه إلى مهرجان السجق الدولي، والذي تنظمه مدينتك أيضًا في نفس عطلة نهاية الأسبوع. (لأجل تأكيد الارتباك، دعنا نفترض إمكانية ارتداء المشاركين في مهرجان السجق سراويل رياضية).

هنيئًا لك. فإذا استطعت إدراك كيف يمكن لأحد ما أن يلقي نظرة خاطفة على أوزان ركاب حافلة ما فيستنتج أنهم على الأرجح ليسوا في طريقهم إلى خط بداية سباق ماراثون، فإنك الآن تفهم الفكرة الأساسية التي تقوم عليها مبرهنة النهاية المركزية. وما الباقي إلا تجسيد لتفاصيل. وإذا ما فهمت مبرهنة النهاية المركزية، فسيبدو لك أي شكل من أشكال الاستدلال الإحصائي أمرًا بديهيًّا نوعًا ما.

المبدأ الأساسي الذي تقوم عليه مبرهنة النهاية المركزية يتلخص في كون العينات الكبيرة المنتقاة بشكل صحيح، ممثِّلة للمجتمع الذي انتُقيت منه. من الطبيعي أن يكون هناك تباين من عينة لأخرى (فمثلًا، سيكون في كل حافلة متجهة إلى خط بداية الماراثون اختلاف طفيف عن غيره من ناحية توليفات جنسيات الركاب)، لكن احتمالاتِ أن تنحرف أيُّ عينةٍ كانت انحرافًا شديدًا عن المجتمع الذي تمثله، احتمالاتٌ ضعيفة جدًّا. هذا المنطق هو الذي دفعك لأن تُصْدر ذلك الحكم السريع ما إن دلفتَ إلى الحافلة المعطلة ورأيت أحجام الركاب الذين يستقلونها. هنالك الكثير من ضخام البنية يَعْدُون في سباقات الماراثون، ومن المحتمل أن يتواجد مئات ممن تتعدى أوزانهم مائتي رطل في أيٍّ من تلك السباقات. ولهذا فإن احتمالات توزيع كل ذلك العدد من العدَّائين الضخام الجثة بشكل عشوائي على نفس الحافلة، ضعيفة جدًّا جدًّا. هنا تستطيع أن تستخلص بقدر كبير من الثقة أن تلك الحافلة

ليست هي حافلة الماراثون التائهة. نعم، قد لا تكون مصيبًا في هذا، لكن الاحتمالات تخبرنا أنك مصيب على الأرجح.

ذاك هو المنطق الأولي الذي يكمن خلف مبرهنة النهاية المركزية. وعندما نضيف إليه بعض أجراس الخطر وأجهزة الإنذار الإحصائية، حينها نستطيع تقدير مدى أرجحية الصواب والخطأ. فمثلًا قد نحسب متوسط أوزان ١٠٠٠٠ عدَّاء في سباق ماراثون ما فنجده ١٥٥ رطلًا، هناك احتمال يقدر بـ ١ من كل ١٠٠ احتمال أن يكون لِعَينة عشوائية مكونة من ٦٠ من هؤلاء العدَّائين (حافلتنا المفقودة) متوسط أوزان يبلغ ٢٢٠ رطلًا أو أكثر. دعنا الآن نركز على منطق الموضوع، أما الحسابات فسيتَّسع لها وقتنا لاحقًا. تُمكننا مبرهنة النهاية المركزية من أن نقوم بالاستدلالات التالية، وكلها سوف نستكشفها بشكل أكثر عمقًا في الفصل التالي.

١- إذا ما تحصلنا على معلومات مفصلة عن مجتمعٍ ما، فحينها سيتسنى لنا أن نخرج باستدلالات قوية عن أي عينة منتقاة من ذلك المجتمع. فمثلًا لنفترض أن مدير مدرسةٍ ما توافرت لديه معلومات مفصلة عن نتائج الاختبارات الموحدة التي خاضها جميع تلاميذ مدرسته تلك (المتوسط والانحراف المعياري، وهلم جرًّا). ذلك مجتمع ذو صلة. والآن لنفترض أن مسئول التعليم في المنطقة التعليمية سوف يأتي الأسبوع القادم لكي يخضع ١٠٠ من تلاميذ مدرستك منتقون عشوائيًّا لاختبار موحد مماثل. سوف يُستخدم أداء هؤلاء التلاميذ المائة في هذا الاختبار بصفتهم عينة لتقييم أداء المدرسة العام.

إلى أي حد هو واثق ذلك المدير من أن أداء هؤلاء التلاميذ المائة سوف يعكس أداء جميع تلاميذ المدرسة في الاختبارات الموحدة المشابهة؟ إلى حد كبير. طبقًا لمبرهنة النهاية المركزية لن ينحرف معدل نتائج اختبارات عينة التلاميذ المائة بشكل حاد عن معدل نتائج اختبارات المدرسة ككل.

٢- إذا ما تحصلنا على معلومات مفصلة عن عينة منتقاة بطريقة صحيحة (المتوسط والانحراف المعياري) فحينها سيتسنى لنا أن نقوم باستدلالات دقيقة بشكل مذهل عن المجتمع الذي انتُقيت منه تلك العينة. وتلك الفكرة تمضي في الاتجاه المقابل للفكرة السابقة، واضعةً إيانا محل مسئول المنطقة التعليمية الذي يقيِّم مدارس متعددة في المنطقة. وعلى النقيض من مدير المدرسة، فإن هذا المسئول لا يمتلك بيانات الاختبارات الموحدة التي يمتلكها المدير (أو لا يثق بها)، وهي البيانات التي تخص جميع تلاميذ تلك المدرسة، والتي هي المجتمع ذو الصلة، وإنما سوف يقوم بوضع اختبار مشابه من عنده يخضع له عينات عشوائية من كل مدرسة قوام الواحدة منها ١٠٠ تلميذ.

هل يمكن لهذا المسئول أن يتيقن بشكل منطقي من أن الأداء الكلي للمدرسة يمكن قياسه قياسًا منصفًا من خلال نتائج اختبارات عينة قوامها مائة تلميذ؟ نعم، فمبرهنة النهاية المركزية تخبرنا بأن العينة الكبيرة لن تنحرف انحرافًا حادًّا عن المجتمع الذي تمثله – وهو ما يعني أن النتائج التي نخرج بها من العينة (نتائج التلاميذ المائة المختارين بشكل عشوائي) هي ممثل جيد لنتائج المجتمع بأسره (جميع تلاميذ تلك المدرسة). بالطبع تلك هي الآلية التي تعمل بها استطلاعات الرأي. فاستطلاع رأي صحيح منهجيًّا، عينته مكونة من ١٢٠٠ أمريكي، قادر على إخبارنا الكثير عما يفكر به الأمريكيون.

فكِّر في الأمر بهذه الطريقة: إذا كانت النقطة الأولى صحيحة، فإن النقطة الثانية لا بد أن تكون صحيحة كذلك، والعكس صحيح. فإذا ماثلَت العينةُ المجتمعَ الذي انتُقيت منه، فلا بد لذلك المجتمع أن يماثل العينة المنتقاة منه. (فإذا شابه الأبناء آباءهم، فلا بد أن يشابه الآباء أبناءهم كذلك).

٣- إذا ما تحصلنا على بيانات تصف عينة محددة، وبيانات أخرى تصف مجتمعًا بعينه، فحينها نستطيع أن نستنتج إذا ما كانت تلك العينة تتسق مع عينة أخرى قابلة لأن تُنتقى من نفس المجتمع أو لا تتسق. وهي الفكرة التي يعبر عنها مثال الحافلة التائهة الذي أوردناه في بداية الفصل. إننا نعلم متوسط وزن المشاركين في سباق الماراثون (نوعًا ما). ونعلم أيضًا متوسط وزن ركاب تلك الحافلة المعطلة (نوعًا ما). تُمكننا مبرهنة النهاية المركزية من حساب احتمالات كون عينة محددة (مستقلي الحافلة السِّمَان) قد تم انتقاؤهم من مجتمع معين (مضمار الماراثون). إذا جاءت الاحتمالات ضعيفة، فحينها نستطيع أن نستخلص بقدر كبير من الثقة أن هذه العينة لم تُنتقَ من المجتمع محل البحث (كأن لا يبدو مستقلو الحافلة أبدًا كمجموعة من العدَّائين المتجهين إلى خط بداية السباق).

٤- وأخيرًا، إذا ما علمنا الخصائص التي تنطوي عليها العينتان، فحينها نستطيع أن نخمن إذا ما كانت العينتان يرجَّح أنهما انتُقِيَتَا من مجتمع واحد. لنعُد إلى مثال الحافلة (المغرق في العبثية). صرنا نعرف الآن أن هناك سباق ماراثون جاريًا في المدينة، وهناك أيضًا مهرجان دولي للسجق. لنفترض أن في كلا الحدثين آلاف المشاركين، وكليهما يَستخدمان حافلات مُلِئَت بعينات عشوائية؛ إما بعدَّائي الماراثون، أو بمهاويس السجق. ولنفترض أيضًا أن الحافلتين قد تصادمتا (لقد أقررت لك أن المثال عبثي، فلتواصل القراءة إذن). وبصفتك قائدًا مدنيًّا فقد وصلت إلى موقع الحادث، وقد أُوكِلَت إليك مهمة تحديد إذا ما كانت الحافلتان كلتيهما كانتا متجهتين إلى نفس الحدث (إما مهرجان السجق أو سباق الماراثون). وبمعجزة ما

لم تجد أيًّا من هؤلاء الركاب يتحدث الإنجليزية، لكن المسعفين في موقع الحادث زوَّدوك بمعلومات مفصلة عن أوزان مستقلي الحافلتين جميعًا.

من تلك المعلومات وحدها تستطيع أن تحدد إذا ما كانت الحافلتان متجهتين معًا إلى نفس الحدث أو كل منهما متجهة إلى حدث مختلف. مرة أخرى عليك أن تفكر بشكل منطقي. لنفترض أن متوسط وزن الركاب على إحدى الحافلات هو ١٥٧ رطلًا بانحراف معياري قدره ١١ رطلًا (وهو ما يعني أن عددًا كبيرًا من الركاب يتراوح وزنهم بين ١٤٦ و١٦٨ رطلًا). والآن لنفترض أن ركاب الحافلة الثانية يبلغ متوسط أوزانهم ٢١١ رطلًا بانحراف معياري قدره ٢١ رطلًا (بمعنى أن عددًا كبيرًا من الركاب يزِنون ما بين ١٩٠ رطلًا و٢٣٢ رطلًا). لنَنْسَ المعادلات الرياضية لبرهة ونكتفِ باستخدام المنطق: هل يبدو لك من المحتمل أن يكون ركاب هاتين العينتين قد اختيروا بشكل عشوائي من نفس المجتمع؟

كلا، فما يبدو راجحًا هنا هو أن واحدة من الحافلتين بها عدَّاء والماراثون، وفي الأخرى مهاويس السجق. فبالإضافة إلى الاختلاف الكائن بين متوسطَي أوزان الحافلتين، تستطيع أن ترى كيف أن التباينات في الوزن بين الحافلتين كبيرة جدًّا مقارنةً بتباينات الأوزان الحادثة في داخل الحافلة الواحدة. فأولئك الرفاق في حافلة «النحفاء»، الذين يقع وزنهم على بُعدِ انحرافٍ معياريٍّ واحدٍ من فوق المتوسط، يبلغ وزنهم ١٦٨ رطلًا، وهو الوزن الذي يقل عن وزن الرفاق في الحافلة «الأخرى» الذين يقع وزنهم على بُعدِ انحرافٍ معياريٍّ واحدٍ أسفل المتوسط (١٩٠ رطلًا). تلك إشارة منبهة (إحصائيًّا ومنطقيًّا في نفس الوقت) إلى أن هاتين العينتين قد أُخذتا، على الأرجح، من مجتمعين مختلفين.

إذا كان ما سبق يبدو لك منطقيًّا، فإنك في طريقك لفهم مبرهنة النهاية المركزية بنسبة ٩٣٫٢ *. لكننا بحاجة لأن نمضي لخطوة إضافية كي نتعرف على بعض الأركان التقنية التي يرتكز إليها ذلك المنطق. فمن الجلي أنك وقت أن دَسَسْتَ رأسك داخل الحافلة فرأيت رهطًا من الجثث الضخام ترتدي سراويل رياضية، خامرك «شعور» بأنهم ليسوا عدَّائي ماراثون. تساعدنا مبرهنة النهاية المركزية على المضي إلى ما وراء ذلك الشعور وخلع قدر كبير من الثقة على تلك الخلاصة التي انتهيت إليها.

فمثلًا سوف تمكنني بعض الحسابات الأولية من التوصل إلى أنه في ٩٩ مرة من أصل ١٠٠ سيكون متوسط وزن أي حافلة منتقاة عشوائيًّا يقع داخل نطاق تسعة أرطال من متوسط أوزان كل المشاركين في مضمار السباق. ذاك هو ما يُكسب حدسي ثقلًا إحصائيًّا عندما أعثر على تلك الحافلة المعطوبة. فهؤلاء الركاب يعلو متوسط أوزانهم بواحد وعشرين رطلًا عن متوسط أوزان المشاركين في السباق، وهو أمر لا يحتمل إلا بنسبة ١ إلى ١٠٠. ونتيجة لهذا أستطيع أن أرفض الفرضية التي تقول بأن تلك هي الحافلة المفقودة بنسبة

* لاحظ الاستخدام الحاذق للدقة الزائفة هنا.

ثقة قدرها ٩٩ بالمائة – وهو ما يعني أن عليَّ أن أتوقع أن استدلالي صحيح ٩٩ مرة من كل مائة.

ونعم، تفترض الاحتمالات أنني في المتوسط قد أكون مخطئًا مرة واحدة من كل ١٠٠ مرة.

هذا النوع من التحليل نابع كله من مبرهنة النهاية المركزية، والتي هي من وجهة نظر إحصائية تمتلك كل ما لِليبرون جيمس من قوة وأناقة. تقضي مبرهنة النهاية المركزية بأن متوسطات عينة أي مجتمعٍ كان تتوزَّع توزُّعًا طبيعيًّا نوعًا ما حول متوسط المجتمع كله. تَمهَّل قليلًا بينما نحلحل عقد تلك العبارة.

١- لنفترض أن لدينا مجتمعًا ما، كالمشاركين في الماراثون مثلًا، وأننا مهتمون بأوزان أولئك الأفراد. سيكون لكل عينة منهم، كمن يستقلون كل حافلة، متوسط للأوزان.

٢- إذا ما واصلنا أخذ عينات مكررة، كأن ننتقي عينات عشوائية قوام الواحدة منها ستون عدَّاءً، المرة تلو الأخرى، فسيكون لكل واحدة منها متوسط أوزان. وتلك هي متوسطات أوزان العينات.

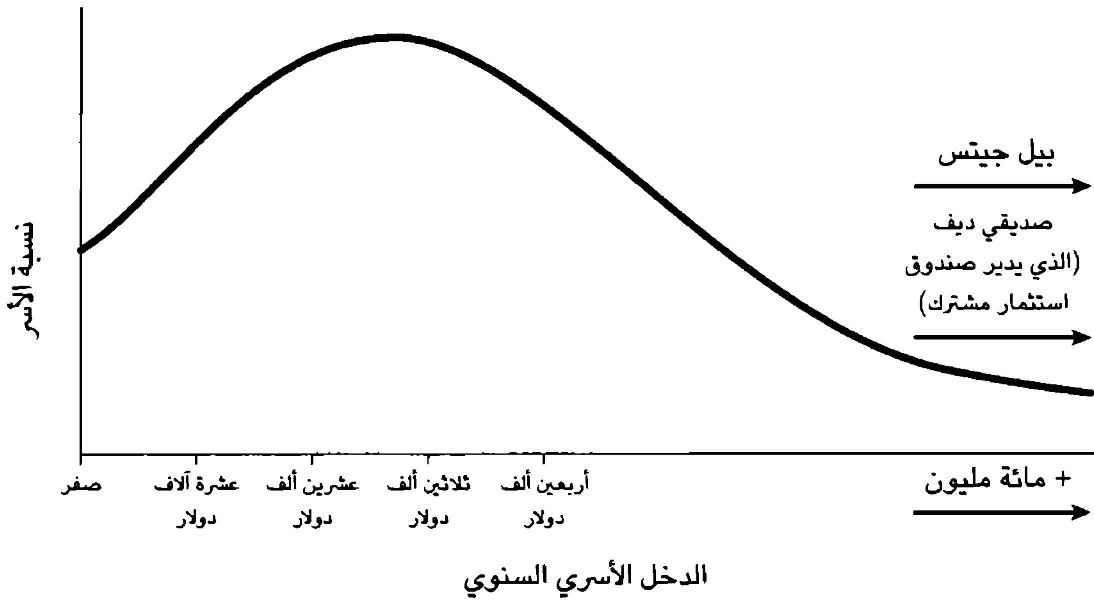
٣- معظم متوسطات أوزان العينات تكون قريبة جدًّا لمتوسط المجتمع كاملًا. بعضها سيكون أعلى، وبعضها سيكون أقل. لكن بعضها، عن طريق الصدفة المحضة، سيعلو بشكل حاد عن متوسط المجتمع كاملًا، وبعضها سينخفض عنه بشكل حاد. فلتصدح الموسيقى هنا حيث تتجمع الأحداث في تصاعد طاغٍ.

٤- تخبرنا مبرهنة النهاية المركزية بأن متوسط العينة سوف يتوزَّع كتوزُّع طبيعي نوعًا ما حول متوسط المجتمع كله. والتوزُّع الطبيعي، كما قد تذكُر من الفصل الثاني، هو توزُّع يتخذ شكل الجرس (كما في توزُّع أطوال الرجال) وفيه تقبع ٦٨ بالمائة من المشاهدات في داخل نطاق انحراف معياري واحد من المتوسط، وتقبع ٩٥ بالمائة من المشاهدات في نطاق انحرافين معياريين منه، وهلم جرًّا.

٥- كل ذلك سوف يكون حقيقيًّا بصرف النظر عن الشكل الذي يتخذه توزيع المجتمع بأكمله. فليس من اللازم أن يكون للمجتمع الذي اختيرت العينة من بينه توزُّعٌ طبيعيٌّ كي تتوزَّع متوسطات عيناته بشكل طبيعي.

لنتأمل بعض البيانات الواقعية كتوزيع دخول الأسر في الولايات المتحدة الأمريكية. لا تتوزع الدخول الأسرية في الولايات المتحدة الأمريكية توزُّعًا طبيعيًّا، وإنما تميل للانحراف يمينًا. فلا يمكن لأسرة أن تجني أقل من ٠ دولار في السنة؛ لذا فلا بد أن يكون هذا هو الحد الأدنى للتوزُّع. لكن في الوقت نفسه قد تجني بعض الأسر الصغيرة دخولًا سنوية هائلة قد تصل إلى مئات الملايين أو حتى مليارات الدولارات في بعض الحالات. ونتيجة لهذا نتوقع أن

يكون توزُّع الدخول الأسرية ذو ذيل طويل متجه ناحية اليمين – شيء على غرار هذا الشكل:



يقدر وسيط الدخل الأسري في الولايات المتحدة بنحو ٥١٩٠٠ دولار، بينما يقدر المتوسط بـ ٧٠٩٠٠ دولار[١]. (يقوم أمثال بيل جيتس بسحب متوسط الدخل الأسري نحو اليمين تمامًا، كما فعل عندما دلف إلى تلك الحانة في الفصل الثاني). والآن لنفترض أننا أخذنا عينة عشوائية مؤلفة من ١٠٠٠ أسرة أمريكية وجمعنا معلومات عن الدخل السنوي الأسري. إذا اعتمدنا على المعلومات التي ذكرناها لتوِّنا وعلى مبرهنة النهاية المركزية، فماذا نستنتج عن هذه العينة؟

سنستنتج الكثير من الأمور، كما اتضح. بادئ ذي بدء سيكون متوسط المجتمع الذي أُخذت منه العينة هو أفضل تخميناتنا لمتوسط العينة نفسها. فالمقصد العام من أخذ تلك العينات الممثلة هو أن تكون مشابهة للمجتمع الذي انتُقيت منه. فعينة مثل تلك تنتقى بشكل مناسب سوف تبدو في المتوسط كأمريكا. ففيها نجد مديرين لصناديق استثمار مشتركة، وأناسًا مشردين بلا مأوًى، ورجال شرطة، وممثلين لجميع البشر – يتناسب عدد كل منهم في العينة إلى حدٍّ ما مع عددهم في المجتمع بأسره. وبناءً عليه نتوقع أن يكون قدر متوسط الدخل الأسري للعينة الممثلة المكونة من ١٠٠٠ أسرة أمريكية، نحو ٧٠٩٠٠ دولار. هل سيكون كذلك بالضبط؟ لا، لكنه لن يختلف عن ذلك الرقم اختلافًا متطرفًا.

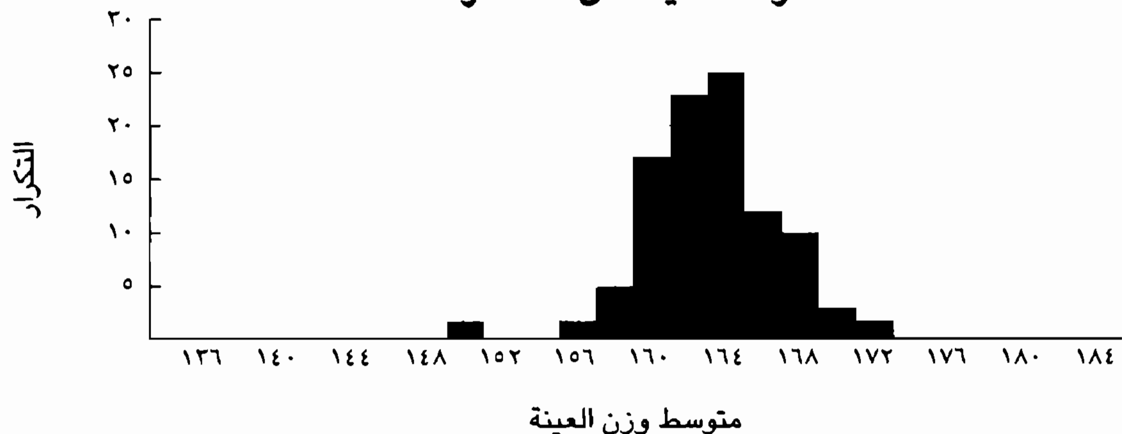
وإذا ما أخذنا عينات متعددة، كلٌّ منها مكون من ١٠٠٠ أسرة أمريكية، نتوقع أن تتحلق متوسطات العينات المختلفة حول متوسط المجتمع كاملًا، والذي هو ٧٠٩٠٠ دولار. قد نتوقع أن ترتفع بعض المتوسطات أو تنخفض عن متوسط المجتمع. هل يمكن أن نجد عينة ما ذات متوسط للدخل الأسري يبلغ ٤٢٧٠٠٠؟ بالطبع يمكن – لكنه أمر مستبعد.

(تذكر أن منهجيتنا في اتخاذ العينات منهجية سليمة، فإننا نجري هذا الاستقصاء في مرآب سيارات بنادي جرينويش الريفي). كما أنه من المستبعد جدًّا أن يكون لإحدى العينات المنتقاة بشكل جيد متوسطُ دخلٍ أسريٍّ يبلغ ٨٠٠٠ دولار.

ما كل ذلك إلا منطق أساسي. ومبرهنة النهاية المركزية تمكننا من المضي في خطوة إضافية لوصف التوزيع المتوقع لمتوسطات تلك العينات بينما تتحلق حول متوسط المجتمع بأكمله. ولنكون أكثر تحديدًا، نقول إن متوسطات العينة سوف تشكل توزُّعًا طبيعيًّا حول متوسط المجتمع، والذي هو في هذه الحالة ٧٠٩٠٠ دولار. تذكر أن شكل المجتمع لا يهم. إن توزُّع الدخل الأسري في الولايات المتحدة منحرف كثيرًا، لكن توزيع متوسطات العينات ليس منحرفًا على الإطلاق. وإذا ما أخذنا ١٠٠ عينة مختلفة، كلٌّ منها تحتوي على ١٠٠٠ أسرة أمريكية، وقمنا برسم بياني لنتائجنا، فلنا أن نتوقع أن تتشكل متوسطات تلك العينات في شكل توزيع يتخذ شكل الجرس الذي نألفه حول ٧٠٩٠٠ دولار.

كلما زاد عدد العينات تقارَب توزيعها مع التوزيع الطبيعي. وكلما زاد حجم كل عينة كان ذلك التوزيع أكثر تلاصقًا. لاختبار تلك النتائج دعنا نقُم بتجربة ممتعة باستخدام بيانات حقيقية لأوزان أمريكيين حقيقيين. تجري جامعة ميتشيجن دراسة طولية تسمى «حيوات الأمريكيين المتغيرة»، والتي تتكون من مشاهدات مفصلة لبضعة آلاف من الأمريكيين البالغين، وأوزانهم من بينها. نجد فيها توزيع الوزن منحرفًا قليلًا نحو اليمين لأنه من الأسهل بيولوجيًّا أن يزيد المرء مائة رطل عن متوسط الوزن، على أن يقل مائة رطل عنه. قدر متوسط الوزن لجميع البالغين في تلك الدراسة بـ ١٦٢ رطلًا.

إذا استخدمنا أحد البرامج الإحصائية الأولية، يمكننا أن نوجه الحاسوب كي يأخذ عينة عشوائية قوامها مائة فرد من بيانات تلك الدراسة. بل إننا قد نكرر العملية مرارًا وتكرارًا كي نرى كيف تتسق النتائج مع ما توقعته مبرهنة النهاية المركزية. الرسم التالي يمثل توزيع متوسطات مائة عينة (مقربة إلى أقرب رطل) ولَّدتها بيانات الدراسة بشكل عشوائي.



كلما ازداد عدد العينات وزاد حجم كل عينة منها، تقارَب متوسط العينة من المنحنى الطبيعي. (من واقع الخبرة أقول إن حجم العينة لا بد ألا يقل عن ٣٠ فردًا كي يصدُق توقُّع مبرهنة النهاية المركزية). هذا أمر منطقي، فالعينة الكبيرة أقل عرضة للتأثر بالتباينات العشوائية. وأي عينة تتألف من شخصين قد تنحرف كثيرًا إذا ما كان بين الشخصين شخص ضخم جدًّا أو ضئيل الجسد جدًّا. لكن على النقيض من ذلك، لن تتأثر عينة مكونة من خمسمائة فرد بحفنة من الضخام أو النحاف. مكتبة سُر مَن قرأ

ها نحن الآن قد اقتربنا من تحقيق جميع أحلامنا الإحصائية. فمتوسط العينة موزَّع في منحنًى طبيعي نوعًا ما، كما ذكرنا من قبل. إن التوزيع الطبيعي يستمد قوته من حقيقة أننا ندرك نوعًا ما كم قدر المشاهدات التي تقبع على بُعد انحراف معياري واحد فوق المتوسط أو تحته (٦٨ بالمائة)، وكم قدر المشاهدات التي تقبع على بُعد انحرافين معيارين فوق المتوسط أو تحته (٩٥ بالمائة)، وهلم جرًّا. إنها معلومات قيِّمة.

كنت قد أشرت في موضع سابق من هذا الفصل إلى أنه في استطاعتنا أن نستنتج منطقيًّا أن تلك الحافلة، ذات الركاب الذين يزيد متوسط أوزانهم عن متوسط أوزان كل المشاركين في الماراثون بخمسة وعشرين رطلًا، ليست هي حافلتنا التائهة. ولتقدير كمية تلك البداهة المنطقية – أن نستطيع القول بأن هذا الاستنتاج سيكون صحيحًا ٩٥ بالمائة من الوقت أو ٩٩ بالمائة أو ٩٩٫٩ بالمائة – نحتاج لمفهوم تقني إضافي واحد؛ ألا وهو الخطأ المعياري.

يقيس الخطأ المعياري مدى تشتت متوسطات العينات. إلى أي مدًى نتوقع أن تكون متوسطات العينات متجمعة بتلاصق حول متوسط المجتمع بأكمله؟ لكن هنا تبرز احتمالات حدوث نوع من أنواع الخلط؛ حيث إننا وجدنا أنفسنا أمام مقياسين مختلفين من مقاييس التشتت، ألا وهما الانحراف المعياري والخطأ المعياري. تحتاج لأن تتذكر النقاط التالية كي تستطيع التفرقة بينهما:

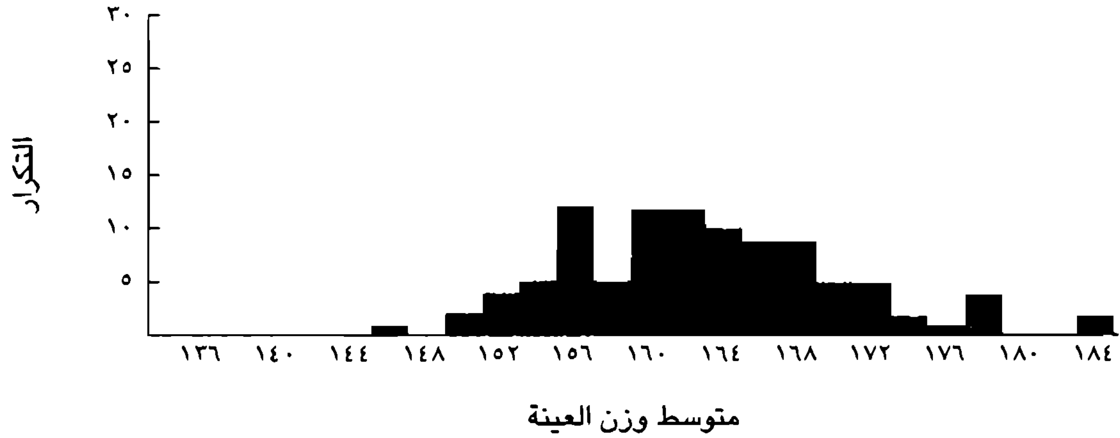
١- يقيس الانحراف المعياري التشتت في المجتمع المدروس. وفي هذه الحالة هو قادر على قياس تشتت الأوزان لدى جميع المشاركين في دراسة فارمنجهام لمرض القلب، أو قياس تشتت أوزان عدَّائي الماراثون حول المتوسط.

٢- يقيس الخطأ المعياري تشتت متوسطات العينات. فإذا ما كررنا سحب عينات قوامها مائة مشارك في دراسة فارمنجهام، فكيف سيبدو تشتت متوسطات تلك العينات؟

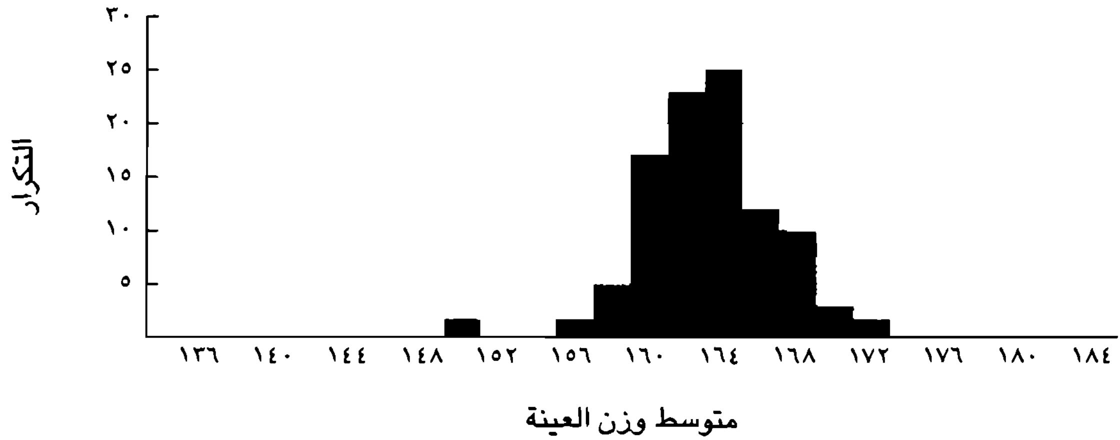
٣- إن ما يربط بين هذين المفهومين هو أن الخطأ المعياري هو الانحراف المعياري لمتوسطات العينات، أليس هذا رائعًا؟

إن الخطأ المعياري الكبير يعني أن متوسطات العينات تتشتت على نطاق واسع حول متوسط المجتمع كاملًا، أما الخطأ المعياري الصغير فيعني أنها متجمعة بشكل متلاصق نسبيًّا. فيما يلي ثلاثة أمثلة حقيقية من دراسة الحيوات المتغيرة.

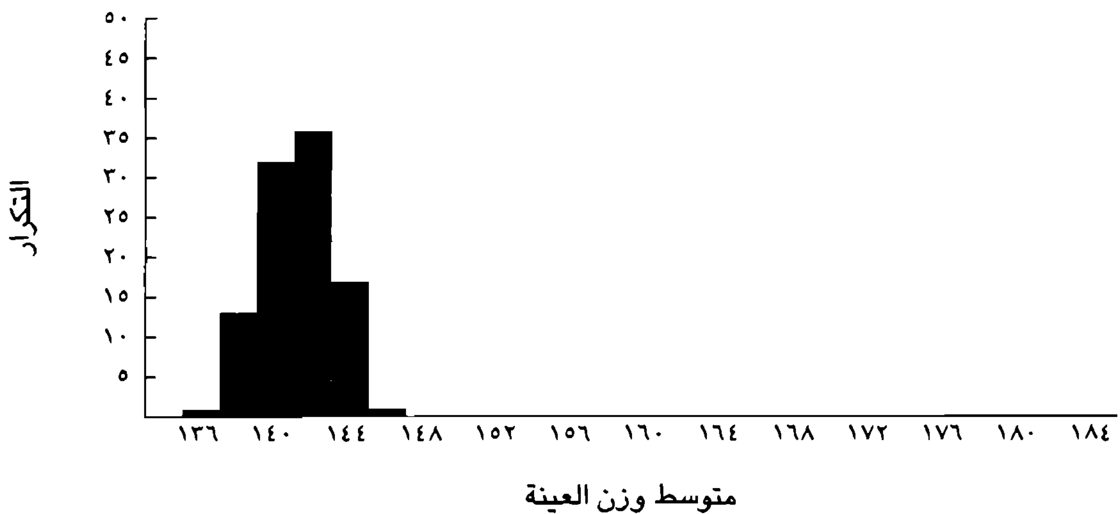
متوسط عينة قوامها ١٠٠ فرد n = ٢٠



متوسط عينة قوامها ١٠٠ فرد n = ١٠٠



متوسط عينة قوامها ١٠٠ من الإناث n = ١٠٠



ذلك التوزيع الثاني ذو العينة الأكبر يتجمع متلاصقًا حول المتوسط أكثر مما يحدث في التوزيع الأول. كلما زاد حجم العينة تراجعت احتمالات انحراف متوسط العينة انحرافًا حادًّا عن متوسط المجتمع. أما المجموعة الأخيرة من متوسطات العينات، فقد سُحبت من مجموعة فرعية من المجتمع، ألا وهي مجموعة النساء المشاركات في الدراسة. وبما أن أوزان النساء في البيانات أقل تشتتًا من أوزان جميع الأشخاص المشاركين في الدراسة، فإن هذا قد يدفعنا للاعتقاد بأن الأوزان في العينات المتخذة من النساء فقط سوف تكون أقل تشتتًا من العينات المتخذة من كامل مجتمع دراسة الحيوات المتغيرة. (هذه العينات تتجمع حول متوسط مجتمع يختلف اختلافًا طفيفًا؛ لأن متوسط أوزان الإناث في الدراسة يختلف عن كامل مجتمع الدراسة).

ذاك النمط الذي رأيته سابقًا حقيقي بشكل عام. فمتوسطات العينات التي تتجمع بتلاصق حول متوسط المجتمع كاملًا بزيادة حجم كل عينة (مثال. كانت متوسطات عينتنا أكثر تلاصقًا عندما كنا نأخذ عينات قوامها ١٠٠ عنها عندما يكون قوامها ٣٠). وتتجمع متوسطات العينات بشكل أقل تلاصقًا حول متوسط المجتمع كاملًا عندما يكون المجتمع كاملًا أكثر تشتتًا. (مثال. متوسطات عيناتنا لكامل مجتمع دراسة الحيوات المتغيرة كانت أكثر تشتتًا من متوسطات العينات المقتصرة على النساء فقط).

إذا ما كنت قد أدركت ذلك المنطق حتى هذه النقطة، فإن معادلة الخطأ المعياري هي كما يلي:

$SE = s / \sqrt{n}$ ؛ حيث s هي الانحراف المعياري للمجتمع الذي اتُّخذت منه العينة، وn هو حجم العينة. حافظ على انتباهك ولا تدعْ مظهر تلك الأحرف يفسد منطقك. سيكون الخطأ المعياري كبيرًا عندما يكبر الانحراف المعياري للتوزيع. والعينات الكبيرة المنتقاة من مجتمع ذي تشتت كبير هي أكثر قابلية لأن تكون أكثر تشتتًا أيضًا، والعينة الكبيرة المنتقاة من مجتمع أكثر تلاصقًا في تجمعه حول المتوسط هي أيضًا أكثر قابلية لأن تتجمع بتلاصق حول المتوسط. وإذا ما كنا لا نزال نبحث في الأوزان، فعلينا أن نتوقع أن يكون الخطأ المعياري للعينة المنتقاة من مجتمع دراسة الحيوات المتغيرة كاملًا أكبر من نظيره لدى العينة المنتقاة فقط من رجال في العشرينيات من عمرهم.

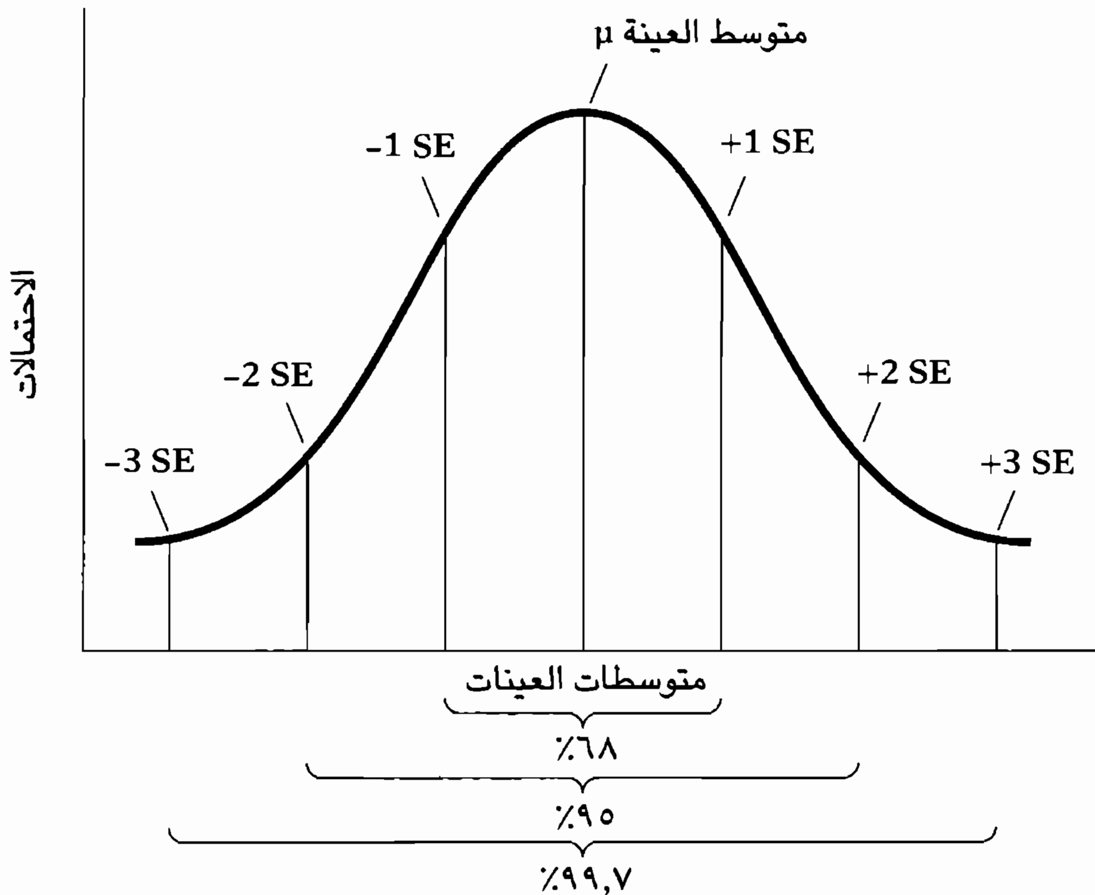
وبالمثل نتوقع أن يصغر الخطأ المعياري بكبر حجم العينة، لأن العينات الكبيرة أقل عرضة للانحراف بفعل القيم المتطرفة؛ ولهذا وجدنا حجم العينة (n) في خانة المقام بالمعادلة. (أما السبب الذي جعلنا نأخذ الجذر التربيعي لحجم العينة n، فسوف نتركه لموضع متقدم من الكتاب، ما يهمنا الآن هو تلك العلاقة الأولية).

إننا في حالة دراسة الحيوات المتغيرة تلك، نحوز الانحراف المعياري للمجتمع كله مسبقًا، لكنها ليست بحالة شائعة. مع العينات الكبيرة نستطيع أن نفترض أن الانحراف

المعياري للعينة قريب نوعًا ما من نظيره لدى المجتمع كاملًا*.

أخيرًا وصلنا إلى ما نرمي إليه من كل ذلك. ولأن متوسطات العينات موزعة بشكل طبيعي (بفضل مبرهنة النهاية المركزية)، يمكننا أن نستفيد من قوة المنحنى الطبيعي. نتوقع أن يقع ما نسبته ٦٨ بالمائة تقريبًا من متوسطات العينات على بُعد خطأ معياري واحد من متوسط العينة، ونحو ٩٥ بالمائة من متوسطات العينات على بُعد خطأين معياريين من متوسط المجتمع، و٩٩,٧ بالمائة تقريبًا على بُعد ثلاثة أخطاء معيارية من متوسط المجتمع.

تكرار توزيع متوسطات العينات



لنعد إلى أحد التنويعات على مثال الحافلة المفقودة السابق، لكننا الآن فقط نستطيع أن نبدل الأرقام لغرض منطقي. (سيظل المثال نفسه على عبثيته، لكنك في الفصل التالي سوف تجد الكثير من الأمثلة الواقعية الأقل عبثية). لنفترض أن القائمين على دراسة

* عندما يحسب الانحراف المعياري للمجتمع كله من خلال عينة صغيرة، نجد المعادلة تتغير قليلًا على هذا النحو: $SE = s / \sqrt{n-1}$ تساعد تلك المعادلة على اعتبار حقيقة أن التشتت في العينات الصغيرة لا يعبر بصدق عن مدى التشتت في المجتمع كاملًا. ليس ذلك ذا صلة كبيرة بالفكرة الأوسع في هذا الفصل.

الحيوات المتغيرة قاموا بدعوة جميع من شاركوا في الدراسة للقاء في بوسطن خلال عطلة نهاية الأسبوع بهدف تجميع البيانات والاحتفال. عُبِّئ المشاركون في الحافلات بشكل عشوائي ونُقِّلوا بين مباني المنشأة التي تجرى بها الاختبارات؛ حيث وُزنوا وقيست أطوالهم ووُخزوا ووُكزوا، وهلم جرًّا. ثم حدثت الصدمة وتاهت إحدى الحافلات وأذيع خبرها في الإذاعة المحلية. وفي نفس الوقت تقريبًا بينما تقود أنت سيارتك عائدًا من مهرجان السجق رأيت حافلة قد اصطدمت بجانب الطريق. يبدو أن الحافلة قد انحرفت عن مسارها كي تتفادى ثعلبًا بريًّا كان يعبر الطريق، مما تسبب في فقدان جميع الركاب لوعيهم، لكن دون أن يصاب أيٌّ منهم بأذًى بليغ. (لأجل المثال أريدهم أن يكونوا غير قادرين على التواصل، لكنني لا أريد أن أفجعك بإصاباتهم الخطيرة). يُبلغك المسعفون بأن متوسط أوزان ركاب الحافلة قدره ١٩٤ رطلًا. كما أن الثعلب الذي فادته الحافلة قد أصيب إصابة غير بالغة وكُسرت إحدى رجليه الخلفيتين.

لحسن حظك أنك تعرف متوسط أوزان كامل مجتمع دراسة الحيوات المتغيرة وتعرف قدر انحرافها المعياري، وكذلك فإنك تمتلك معرفة عملية بمبرهنة النهاية المركزية، وتعرف كذلك كيف تجري الإسعافات الأولية للثعلب. متوسط أوزان المشاركين في الدراسة قدره ١٦٢، وانحرافها المعياري قدره ٣٦. من هذه المعلومات نستطيع أن نحسب الخطأ المعياري لعينة قوامها ٦٢ شخصًا (عدد ركاب الحافلة الفاقدي الوعي) على النحو التالي: $s / \sqrt{62} = 36 / 7.9$ أو ٤,٦.

يبلغ الفارق بين متوسط العينة (١٩٤ رطلًا) ومتوسط المجتمع كاملًا (١٦٢ رطلًا) ٣٢ رطلًا أو ما يزيد عن ثلاثة أخطاء معيارية. نعرف من مبرهنة النهاية المركزية أن ٩٩,٧ بالمائة من جميع متوسطات العينات تقع في نطاق ثلاثة أخطاء معيارية من متوسط المجتمع. وهو الأمر الذي يجعل من المستبعد جدًّا أن تكون تلك الحافلة عينة ممثلة لكامل مجتمع المشاركين في دراسة الحيوات المتغيرة. وطبقًا لواجبك كقائد مدني فقد هاتفت القائمين على الدراسة وأبلغتهم أنها ليست حافلتهم التائهة، على الأرجح، لكنك الآن قادر على أن تدفع لهم بدليل إحصائي لا مجرد «شعور». ترفع تقريرًا إلى أولئك القائمين على الدراسة مفاده أنك ترفض احتمالات كونها الحافلة المنشودة بنسبة ثقة قدرها ٩٩,٧ بالمائة. ولأنك تخاطب باحثين، فهم يدركون ما تتكلم عنه.

يزداد تحليلك تأكيدًا بعدما يجري المسعفون فحوص دم لركاب الحافلة فيكتشفون أن مستويات الكوليسترول في دماء أولئك الركاب تعلو عن متوسط مستوى الكوليسترول في دماء المشاركين في الدراسة بخمسة أخطاء معيارية. من تلك المعلومة نفترض، وسنتأكد فيما بعد، أن أولئك الركاب الذين أصيبوا بالإغماء كانوا مشاركين في مهرجان السجق.

[لكن لتلك الحادثة نهاية سعيدة. فعندما استعاد ركاب الحافلة وعيهم، قدم لهم

القائمون على دراسة الحيوات المتغيرة مشورة تتعلق بمخاطر الإكثار من الدهون المشبعة؛ مما جعلهم يتبعون عادات غذائية أكثر صحية فيما يتعلق بسلامة القلب. وفي الوقت نفسه تمت العناية بالثعلب حتى استعاد عافيته، ثم أودع إحدى محميات الحياة البرية المحلية، وأخيرًا تم إطلاقه مرة أخرى إلى البراري]*.

لقد حاولت في هذا الفصل أن أتمسك بالأساسيات. لا بد أن تضع في اعتبارك أنه كي تعمل مبرهنة النهاية المركزية لا بد أن يكون حجم العينة كبيرًا نسبيًّا (ما يزيد عن الثلاثين من واقع الخبرة). كما أننا نحتاج لِعينة كبيرة نسبيًّا إذا ما كنا سوف نفترض أن الانحراف المعياري لِلْعينة يماثل تقريبًا الانحراف المعياري للمجتمع الذي انتُقيت منه كاملًا. هنالك الكثير من الإجراءات التصحيحية التي يمكن أن تتخذ إذا لم توفَّ تلك الشروط – لكن ما هذا إلا بمثابة القشدة التي تعلو الكعكة (بل وربما كانت السكاكر المنثورة على قشدة الكعكة). و«الصورة الكبرى» هنا بسيطة وقوية للغاية:

١- إذا ما انتقيت عينات عشوائية كبيرة من أي مجتمع كان، تتوزع متوسطات تلك العينات بشكل طبيعي حول متوسط المجتمع (بصرف النظر عن الشكل الذي يبدو عليه توزيع المجتمع).

٢- معظم متوسطات العينات سوف يقع قريبًا بشكل معقول من متوسط المجتمع، والخطأ المعياري هو الذي يعرف معنى «قريبًا بشكل معقول».

٣- تخبرنا مبرهنة النهاية المركزية باحتمالات كون متوسط العينة يقع في نطاق مسافة معينة من متوسط المجتمع. وإنه لمن المستبعد أن يقع متوسط العينة على بُعد أكثر من خطأين معياريين من متوسط المجتمع، ومن الأكثر استبعادًا أن يقع على بُعد ثلاثة أخطاء معيارية أو أكثر من متوسط المجتمع.

٤- كلما قلت، احتمالات ملاحظة ناتجٍ ما عن طريق الصدفة، زادت ثقتنا في تصور وجود عوامل أخرى متحكمة.

ما سبق كان تقريبًا كلَّ شيء يتعلق بالاستدلال الإحصائي. ومبرهنة النهاية المركزية هي التي تجعل معظمه ممكنًا. وإلى أن يربح ليبرون جيمس عددًا من بطولات الرابطة الوطنية لكرة السلة يفوق عدد البطولات التي ربحها مايكل جوردان (ستة)، ستظل مبرهنة النهاية المركزية تفوقه روعة.

---

* ذكر زميلي بجامعة شيكاجو جيم سالي نقدًا مهمًّا لأمثلة الحافلة التائهة. فقد أشار إلى أنه من النادر أن تضل الحافلات طريقها. لذا فإننا إذا ما شرعنا نبحث عن حافلة تائهة فوجدنا إحداها ضالة طريقها كانت أو متعرضة لحادث فإنها، على الأرجح، حافلتنا المنشودة، بصرف النظر عن أوزان راكبيها. إنه محق. (فكر في الأمر بهذه الطريقة، لو أضعت طفلك في أحد المتاجر الكبيرة، ثم أخبرك مدير المتجر أن هناك طفلًا ضاع من أهله موجودٌ عند نافذة الدفع رقم ستة، فسوف تستنتج فورًا أنه طفلك على الأرجح). ومن ثم فإننا سوف نقوم بإضافة عنصر إضافي من عناصر العبثية، ألا وهو تخيل أن الحافلات كثيرًا ما تضل طريقها.

الفصل التاسع

الاستدلال

لماذا ظن أستاذي في مادة الإحصاء أنني قد أكون غششت في الاختبار؟

في ربيع سَنَتِي الأخيرة بالجامعة التحقتُ بصف الإحصاء. لم أكن وقتها مغرمًا بالإحصائيات أو غيرها من المناهج القائمة على الرياضيات، لكنني كنت قد وعدت والدي أنني سألتحق بهذا الصف إذا سمح لي بأن أتغيب عن الدراسة لمدة عشرة أيام كي أذهب في رحلة عائلية إلى الاتحاد السوفييتي. وهكذا فقد كان تعلُّمي للإحصاء مجرد صفقة في مقابل رحلة للاتحاد السوفييتي. وهي الصفقة التي اتضح لي فيما بعد أنها رابحة جدًّا؛ وذلك لأنني أحببت الإحصاء أكثر مما كنت أظن، ولأنها سمحت لي بأن أزور الاتحاد السوفييتي في ربيع عام ١٩٨٨ ومن كان يعرف أن تلك البلاد لم تكن لتظل في شكلها الشيوعي لوقت طويل بعدها؟

هذه القصة هي في الواقع ذات صلة بهذا الفصل، وأقصد هنا أنني لم أكن مكرِّسًا قدرًا كبيرًا من مجهودي لصف الإحصاء خلال ذلك الفصل الدراسي كما كان يفترض بي. فقد كنت مشغولًا، ضمن مسئوليات أخرى، بكتابة أطروحة إلزامية يحين موعد تسليمها في منتصف الفصل الدراسي. كنا نعطى اختبارات سريعة في صف الإحصاء، وكنت أتجاهل الكثير منها أو أرسب فيه. وفي اختبار منتصف الفصل الدراسي أدَّيت أداءً مقبولًا – بالمعنى الحرفي للكلمة. لكن قبل انتهاء الفصل الدراسي بأسبوعين حدث أمران. أولهما أنني انتهيت من أطروحتي؛ مما وفَّر لي وقت فراغ كبيرًا جدًّا. وثانيهما أنني أدركت أن الإحصاء لا يقترب حتى من الصعوبة التي تصورتها. بدأت أذاكر كتاب الإحصاء وأعوض ما فاتني في بداية الصف. ثم حدث أن حصلت على تقدير ممتاز في الاختبار النهائي.

هنا قام أستاذ مادة الإحصاء – الذي نسيت اسمه منذ زمن طويل – باستدعائي إلى مكتبه. لا أذكر ما قاله بالضبط، لكنه كان كلامًا على غرار: «لقد أبليتَ بلاءً حسنًا

في الاختبار النهائي خيرًا مما فعلت في اختبار منتصف الفصل الدراسي». لم يكن ذلك الاستدعاء بغرض الاحتفاء بحسن صنيعي في الاختبار النهائي. وإنما كان ينطوي على اتهام ضمني (غير ظاهر) حيث كان يسعى لأن يجعلني أخبره بسبب هذا التحول الحسن عما حدث في اختبار منتصف العام. باختصار كان الرجل يشك في أنني قد غششت في الاختبار. والآن بعد أن امتهنت التدريس لسنوات طوال، فقد تفهمت تفكيره وقتها. ففي كل صف أدرسه أجد ارتباطًا مدهشًا بين أداء كل طالب في اختبار منتصف العام ونظيره في الاختبار النهائي. فمن المستبعد جدًّا أن تأتي نتيجة الطالب دون المتوسط في اختبار منتصف العام ثم تعلو حتى تقترب من القمة في الاختبار النهائي.

وضحت له أنني بعد أن انتهيت من أطروحتي، بدأت أحمل صف الإحصاء محمل الجد (عن طريق أمور كقراءة الفصول المقررة من المراجع الأكاديمية وإنجاز الفروض المنزلية). هنا بدت عليه علامات الرضا إزاء توضيحي ذلك، وغادرت أنا مكتبه ولا أزال منزعجًا من ذلك الاتهام المبطن.

صدق أو لا تصدق أن هذه الحكاية تجسد كل ما تحتاج لمعرفته عن الاستدلال الإحصائي، شاملًا مَواطن قوته ونقاط ضعفه. لا تستطيع الإحصائيات أن تثبت أي شيء عن يقين، وإنما يَستمد الاستدلال الإحصائي قوَّتَه من ملاحظة الأنماط أو النواتج، ومن ثم استخدام الاحتمالات لتحديد التفسيرات الأكثر أرجحية لهذه النواتج. لنفترض أن مقامرًا غريبًا عن البلدة قد أتى وعرض عليك الرهان التالي: إذا ما رمى هو النرد مرة واحدة فأصاب رقم ستة يربح حينها ١٠٠٠ دولار، أما إذا أصاب أي رقم آخر تربح أنت ٥٠٠ دولار – هو رهان ممتاز من وجهة نظرك. لكن ما حدث أنه أصاب الرقم ستة، عشر مرات متتالية سالبًا إياك ١٠٠٠٠ دولار.

واحد من التفسيرات المحتملة لهذا أنه رجل محظوظ، لكن هناك تفسيرًا آخر، ألا وهو أنه قد غشك بطريقة ما. فاحتمالات إصابة الرقم ستة، عشر مرات متتالية بنرد غير معيوب هي ١ إلى ٦٠ مليونًا. لا تستطيع أن تثبت غشه، لكن من حقك على الأقل أن تفحص النرد.

بالطبع ليس الاحتمال الأرجح هو الاحتمال الصحيح دائمًا. فالأمور الشديدة الندرة تحدث. لندا كوبر هي امرأة تعيش في ساوث كارولينا وقد ضربها البرق أربعة مرات[1] (تقدر الإدارة الفدرالية للتعامل مع الطوارئ احتمالات أن يصاب أحدهم بضربة برق مرة واحدة بـ ١ إلى ٦٠٠٠٠٠). لا يمكن لشركة التأمين التي تغطي لندا كوبر أن تمتنع عن تعويضها لمجرد أن إصابتها تلك غير ممكنة إحصائيًّا. وبالعودة إلى اختبار الإحصاء الذي خُضتُه في الجامعة، فإن الأستاذ كان لديه سبب منطقي للشك. فقد رأى نمطًا مستبعدًا جدًّا، وتلك هي بالضبط الكيفية التي يضبط بها المفتشون حالات الغش في الاختبارات الموحدة، والتي تضبط بها SEC التجارة الداخلية. لكن النمط المستبعد يصير مستبعدًا

عندما لا ترجحه دلائل إضافية. في موضع لاحق من هذا الفصل سوف نناقش الأخطاء التي يمكن أن تنشأ عندما تقودنا الاحتمالات إلى الاتجاهات الخطأ.

حتى هذه النقطة علينا أن نقدِّر حقيقة أن الاستدلال الإحصائي يستخدم البيانات في الإجابة عن الأسئلة المهمة. هل تسبب الهواتف المحمولة الإصابة بمرض السرطان؟ أرجوك لاحظ أنني لا أدَّعي أن الإحصاء قادر على الإجابة عن مثل تلك الأسئلة بإجابات أكيدة لا لبس فيها، وإنما يخبرنا الاستدلال بما هو راجح وما هو غير راجح. فلا يستطيع الباحثون أن يثبتوا نجاعة عقَّار جديد في علاج القلب حتى لو حازوا بيانات مصدرها تجربة سريرية محكمة بعناية. فرغم كل شيء، من الممكن جدًّا أن يوجد تغايُر عشوائي في نواتج المرضى، الذين هم في مجموعتَي المعالجة والضبط ليست ذات علاقة بذلك العقار الجديد. فإذا ما ظهر تحسن على ٥٣ مريضًا من أصل ١٠٠ تناولوا دواء القلب الجديد مقارنةً بـ ٤٩ مريضًا من أصل ١٠٠ تناولوا العقَّار الوهمي، لا نستطيع حينها أن نستنتج بشكل فوري نجاعة ذلك الدواء. فيمكن تفسير ذلك الناتج بكل سهولة على أنه تغايُر حدث عن طريق الصدفة بين المجموعتين ولم يحدث بفعل الدواء.

لكن لنفترض أن ٩١ من أصل المائة الذين تلقوا العقَّار الجديد ظهرت عليهم علامات التحسن، مقارنة بـ ٤٩ من أصل المائة الذين في مجموعة الضبط. لا يزال من الممكن أن تكون تلك النتيجة الرائعة لا علاقة لها بالعقَّار، وأن يكون سببها هو كون أفراد مجموعة المعالجة محظوظين أو مقاومين للمرض. لكن هذا التفسير الآن قد قلَّت احتمالاته جدًّا. لكي نعبِّر عن الأمر بِلُغة الاستدلال الإحصائي الرسمية نقول إن الباحثين غالبًا ما سوف يستخلصون ما يلي: (١) إذا لم يكن لهذا العقَّار التجريبي أي أثر فلن نرى، غالبًا، أي قدر من التغاير في النواتج بين أولئك الذين تلقوا ذلك العقَّار وأولئك الذين تلقوا العقَّار الوهمي. (٢) ومن ثم فإنه من المستبعد جدًّا ألا يكون للعقَّار أية آثار إيجابية. (٣) التفسير البديل – والأكثر احتمالًا – لنمط البيانات الملاحظ ذلك هو أن ذلك العقَّار له أثر إيجابي.

إن الاستدلال الإحصائي هو عملية تُكلِّمُنا البياناتُ من خلالها، وتتيح لنا أن نستخلص منها نتائج ذات معنًى. ذاك هو المكسب. فليس الهدف من الإحصاء إجراء عدد لا يحصى من الحسابات الرياضية المعقدة، وإنما يهدف إلى أن يجعلنا نفهم ظواهر اجتماعية ذات معنًى. فما الاستدلال الإحصائي إلا زواج بين مفهومين كنا قد ناقشناهما سابقًا، ألا وهما البيانات والاحتمالات (مع بعض العون من مبرهنة النهاية المركزية). في هذا الفصل قد اتخذت طريقًا منهجيًّا مختصرًا مهمًّا، ففي كل الأمثلة التي سأوردها سنفترض أن العينات التي نعمل عليها كلها عينات كبيرة منتقاة بصورة صحيحة. وهذا الافتراض يعني أن مبرهنة النهاية المركزية تنطبق على المثال، وأن متوسط العينة وانحرافها المعياري متطابقان مع المتوسط والانحراف المعياري لكامل المجتمع الذي اتُّخذت العينة

منه. وهذان الأمران كلاهما سوف يجعلان حساباتنا أكثر سهولة.

لا يقوم الاستدلال الإحصائي على ذلك الافتراض المُبسَّط وإنما يقوم على الإجراءات التصحيحية المنهجية المتنوعة التي نتعامل بها مع العينات الصغيرة أو البيانات المعيبة التي دائمًا ما تحول بيننا وبين إدراك الصورة الكبيرة. غرضنا هنا هو أن نعرض لقدرة الاستدلال الإحصائي وشرح كيفية عمله. وإذا ما أدركت هذا فسيسهل عليك حينها أن تتغلب على ما فيه من تعقيدات.

واحدة من أكثر أدوات الاستدلال الإحصائي شيوعًا هي اختبار الفرضيات. والواقع أنني قد وضحت لك هذا المفهوم لكن دون أن أستخدم مصطلحاته المنمقة. كما أشرنا سابقًا، فإن الإحصائيات وحدها لا تستطيع أن تثبت أي شيء، وإنما نستخدم الاستدلال الإحصائي وحده كي نقبل أو نرفض بعض التفسيرات استنادًا إلى أرجحيتها النسبية. ولمزيد من التحديد نقول إن أي استدلال إحصائي يبدأ بفرضية منعدمة صريحة أو مبطنة. تلك هي فرضيتنا الأولية التي قد تدحض أو لا تدحض استنادًا إلى ما سوف يلي من تحليل إحصائي. فإذا ما رفضنا فرضية منعدمة فسيعني هذا أننا نَقبل فرضية بديلة أكثر اتساقًا مع البيانات التي لاحظناها. فمثلًا إذا كان الافتراض الأولي أو الفرضية المنعدمة لدى إحدى المحاكم هو أن المدعى عليه بريء. هنا تتمثل وظيفة الادعاء في إقناع القاضي أو هيئة المحلفين برفض هذا الافتراض وقبول الفرضية البديلة، والتي تنص على أن المدعى عليه مذنب. وإذا نظرنا إلى الأمر نظرة منطقية لوجدنا أن الفرضية البديلة ما هي إلا خلاصة لا بد أن تكون صحيحة إذا دحضت تلك الفرضية المنعدمة. لنأخذ في الاعتبار بعض الأمثلة.

الفرضية المنعدمة: هذا العقَّار التجريبي الجديد ليس أكثر فعالية في الوقاية من الملاريا من العقَّار الوهمي.

الفرضية البديلة: هذا العقَّار التجريبي الجديد قادر على المساعدة في الوقاية من الملاريا.

البيانات: أن تختار مجموعة بشكل عشوائي لتلقِّي العقَّار التجريبي، وتختار مجموعة ضابطة أخرى لتتلقى العلاج الوهمي. بنهاية فترة معينة من الزمن صار لدى المجموعة التي تلقَّت العلاج التجريبي حالات ملاريا أقل بكثير من التي صارت لدى المجموعة الضابطة. لو لم يكن للعقَّار التجريبي أي أثر طبي فعال، لكان هذا الناتج مستبعدًا جدًّا. ونتيجة لهذا نستطيع أن نستبعد الفرضية المنعدمة التي تقول بأن الدواء ليس له من أثر (يفوق أثر العقَّار الوهمي)، ونتقبل البديل المنطقي الذي هو الفرضية البديلة، والتي تنص على أن العقَّار التجريبي البديل قادر على المساعدة في الوقاية من الملاريا.

هذه المقاربة المنهجية من الغرابة بمكان أننا نحتاج لمثال آخر عليها. مرة أخرى لاحظ أن الفرضيتين المنعدمة والبديلة تتكاملان منطقيًّا. فإذا ما صحت إحداهما بطلت

الأخرى. أو إذا ما رفضنا نتيجة إحداهما فلا بد أن نقبل نتيجة الأخرى.

الفرضية المنعدمة: علاج إدمان المخدرات الذي يتلقاه السجناء لا يقلل من فرصة تعرُّضهم للاعتقال بعد إطلاق سراحهم.

الفرضية البديلة: علاج إدمان المخدرات الذي يتلقاه السجناء سوف يقلل من احتمالات تعرُّضهم للاعتقال بعد إطلاق سراحهم.

البيانات (الافتراضية): تم توزيع السجناء بشكل عشوائي على مجموعتين، فتلقَّت مجموعة «المعالجة» علاجًا لإدمان المخدرات لم تتلقَّه المجموعة الضابطة. (تلك واحدة من الفرص السعيدة التي تتلقى فيها مجموعة المعالجِ معالجةً حقيقية) وبعد انقضاء خمسة أعوام صار لدى المجموعتين كليهما معدلات متشابهة لمن اعتقلوا مرة أخرى بعد إطلاق سراحهم. في هذه الحالة لا نستطيع أن نرفض الفرضية المنعدمة*. لم تمنحنا البيانات سببًا لنبذ افتراضنا الأولي بأن علاج إدمان المخدرات ليس بالأداة الفعالة التي تَحُول بين المحكومين السابقين والعودة إلى السجن.

قد يبدو ذلك الكلام مضادًّا للمنطق، لكن الباحثين عادةً ما يخلقون الفرضية المنعدمة أملًا في أن يستطيعوا دحضها. في المثالين السابقين كليهما كان «نجاح» البحث (إيجاد عقَّار جديد لعلاج الملاريا أو تقليل فرص الانتكاس والعودة للمخدرات) ينطوي على رفض جميع الفرضيات المنعدمة. لقد أتاحت البيانات هذا الأمر في حالة واحدة فقط من هاتين الحالتين (عقَّار الملاريا).

في قاعات المحاكم تكون النقطة التي عند الوصول إليها ترفض فرضية البراءة هي التقييم النوعي الحاكم بأن المدعى عليه «مذنب دونما أي شك منطقي». أما عن معنى هذا فهو يترك للقاضي أو لهيئة المحلفين كي يحددوه. والإحصاء يقوم بتعزيز نفس تلك الفكرة الأولية، لكن يتم تعريف عبارة «مذنب دونما أي شك منطقي» بشكل كمي هذه المرة. عادة ما يطرح الباحثون سؤالًا مفاده أنه إذا صحت الفرضية المنعدمة، فإلى أي مدى من المرجح أن نلاحظ نحن النمط الذي تنتظم فيه البيانات عن طريق المصادفة؟ للتقريب نلجأ إلى مثالنا الذي ألَّفناه فنقول إن الباحثين الطِّبيين قد يتساءلوا قائلين إذا لم يكن لذلك

* من وجهة نظر علم المعاني، لم نثبت نحن صحة الفرضية المنعدمة (أي إن علاج إدمان المخدرات لا أثر له). فقد تثبت فعاليتها الكبيرة لدى مجموعة أخرى من السجناء. أو ربما اعتقل المزيد من السجناء الذين وزعوا على مجموعة المعالجة إذا لم يتلقوا العلاج. في أي حالة كانت، إننا قد فشلنا في رفض الفرضية المنعدمة استنادًا إلى البيانات التي جمعناها. هناك فرق مشابه بين «الفشل في رفض» الفرضية المنعدمة وقبول الفرضية المنعدمة. فقط لأن دراسة واحدة لم تستطع تفنيد كون علاج إدمان المخدرات ليس له أثر (نعم، هذا نفي مزدوج)، فلا يعني هذا أن علينا أن نقبل كون علاج إدمان المخدرات غير ذي نفع. هنا نجد فارقًا إحصائيًّا ذا معنًى. بعد أن ذكرنا هذا لا بد أن نقول إن الأبحاث عادة ما تصمَّم لإعلام واضعي السياسات، ومسئولي السجون، وأولئك الذين يقررون كيف توزع الموارد بأنهم قد يقبلون فرضية كون علاج إدمان المخدرات غير ذي فعالية إلى أن يقتنعوا بالعكس. وهنا كما في الكثير من مجالات الإحصاء يكتسب حسن التقدير أهمية.

العقَّار التجريبي أي تأثير على مرض القلب (الفرضية المنعدمة)، فإلى أي مدًى من المرجح أن تظهر علامات التحسن على ٩١ مريضًا من أصل ١٠٠ قد تلقَّوا العقَّار بالفعل مقارنةً بـ ٤٩ مريضًا من أصل ١٠٠ تلقَّوا العقَّار الوهمي؟ فإذا ما أفادت البيانات بأن الفرضية المنعدمة مستبعدة التحقق جدًّا – كما في هذا المثال الطبي – يتعين علينا حينها أن نرفضها ونقبل الفرضية البديلة (والتي تفيد بأن العقَّار مفيد في علاج مرض القلب).

في ذات الإطار دعنا نَعُدْ إلى فضيحة الغش في الاختبارات الموحدة بأطلانطا، والتي أشرنا لها مرارًا في هذا الكتاب. لقد لفتت نتائج الاختبارات بأطلانطا الانتباه في البداية بسبب زيادة عدد الإجابات التي عُدِّلت فتحولت من الخطأ إلى الصواب. بالطبع لا يفتأ الطلبة الذين يخضعون للاختبارات الموحدة يمحون إجابتهم ويعدلونها طوال الوقت. وبعض من يفعلون ذلك يحالفهم الحظ في تعديل إجاباتهم تلك دون أن ينطوي هذا التعديل على أي غش بالضرورة. لذلك السبب تقضي الفرضية المنعدمة بأن أي نتائج لاختبارات موحدة، لأي منطقة تعليمية كانت، هي نتائج شرعية، وأن أي أنماط غير معتادة تظهر في تعديل الطلاب لإجاباتهم ما هي إلا وليدة للصدفة. وبالطبع لا أحد يريد أن يعاقب التلاميذ ولا مديري المدارس لمجرد أن «شطرًا كبيرًا غير معتاد» من التلاميذ قد وفقوا في تلك التعديلات التي أجروها في أوراق إجاباتهم في الدقائق الأخيرة من زمن اختبار الولاية الحرج.

لكن عبارة «شطرًا كبيرًا غير معتاد» لا تكاد تقترب من وصف ما حدث في أطلانطا. فقد كانت هناك صفوف كاملة بلغ فيها معدل تعديلات الإجابات من عشرين إلى خمسين انحرافًا معياريًّا فوق المعدل المعتاد للولاية. (لتوضيح الصورة تَذكَّرْ أن معظم المشاهدات تقع في التوزيع في نطاق انحرافين معياريين من المتوسط). فما هو مدى احتمالية أن يكون تلاميذ أطلانطا قد تصادف أنهم محوا عددًا هائلًا من الإجابات، واستبدلوا بها إجابات صحيحة فقط عن طريق الصدفة؟ شبَّه المسئولُ الذي قام بتحليل تلك البيانات احتمالاتِ حدوث ذلك النمط الذي جرى في أطلانطا، مع افتراض عدم حدوث غش، باحتمالات تصادف حضور ٧٠٠٠٠ شخص بين جمهور مباراة كرة قدم تقام في ملعب جورجيا دوم، وكلهم يتجاوزون السبع أقدام طولًا[٢]. هل يمكن أن يحدث ذلك؟ نعم. هل من المرجح حدوثه؟ لا ليس كثيرًا.

لكن رغم ذلك لم يكن لمسئولي جورجيا أن يُدينوا أي أحد بذلك الجرم، تمامًا كما لم يستطع أستاذي (ولم يكن يجدر به) أن يفصلني من الجامعة لأن علاماتي في الاختبار النهائي لمادة الإحصاء لم تكن تتسق مع علاماتي في اختبار منتصف الفصل الدراسي. لم يستطع مسئولو أطلانطا أن يثبتوا ما يجري من غش. لكنهم رغم ذلك قادرون على رفض تلك الفرضية المنعدمة التي تقضي بشرعية تلك النتائج. ويمكنهم أن يفعلوا ذلك «بدرجة

عالية من الثقة»؛ بمعنى أن ذلك النمط الذي لاحظوه يكاد يستحيل حدوثه بين الخاضعين للاختبار من التلاميذ العاديين. لذلك فقد قبلوا الفرضية البديلة، والتي تقضي بحدوث شيء مريب في الموضوع. (لكني أظنهم استخدموا مصطلحات أكثر رسمية). ولقد كشف التحقيق الذي تلا ذلك عن «تعديلات إجابات مريبة تكاد تقول خذوني». فقد وردت تقارير عن معلمين قاموا بتغيير بعض الإجابات وأملوا بعض الإجابات وسمحوا لتلاميذ ذوي علامات ضعيفة أن ينقلوا من إجابات تلاميذ ذوي علامات ممتازة، بل وحتى أن يشيروا لبعض التلاميذ على الإجابات الصحيحة بينما يقفون قرب مقاعدهم وقت الاختبار. أما أكثر وقائع الغش فجاجة فكانت وقت أن أقام مجموعة من المعلمين حفل بيتزا في عطلة نهاية الأسبوع، قاموا خلاله بتصفح أوراق إجابات التلاميذ معدلين الإجابات الخطأ إلى أخرى صحيحة.

في مثال أطلانطا يمكننا أن نرفض الفرضية المنعدمة التي «تنفي الغش»؛ لأن نمط إجابات الاختبارات يغدو مستبعدًا جدًّا في غياب انتهاك القواعد. لكن إلى أي مدًى ينبغي على الفرضية المنعدمة أن تكون غير قابلة للتصديق كي يتسنى لنا أن نرفضها ونلجأ إلى التفسير البديل؟

من أكثر الحدود التي يستعملها الباحثون لدحض هذه الفرضية المنعدمة شيوعًا هي نسبة ٥ بالمائة، وعادة ما تكتب في شكل عدد عشري على هذا النمط ٠٫٠٥. تعرف هذه الاحتمالات باسم مستوى الدلالة، وتمثل الحد الأعلى لأرجحية ملاحظة نمطٍ ما من البيانات إذا ما صحت الفرضية المنعدمة. ابقَ معي للحظات، فالأمر ليس بهذا التعقيد.

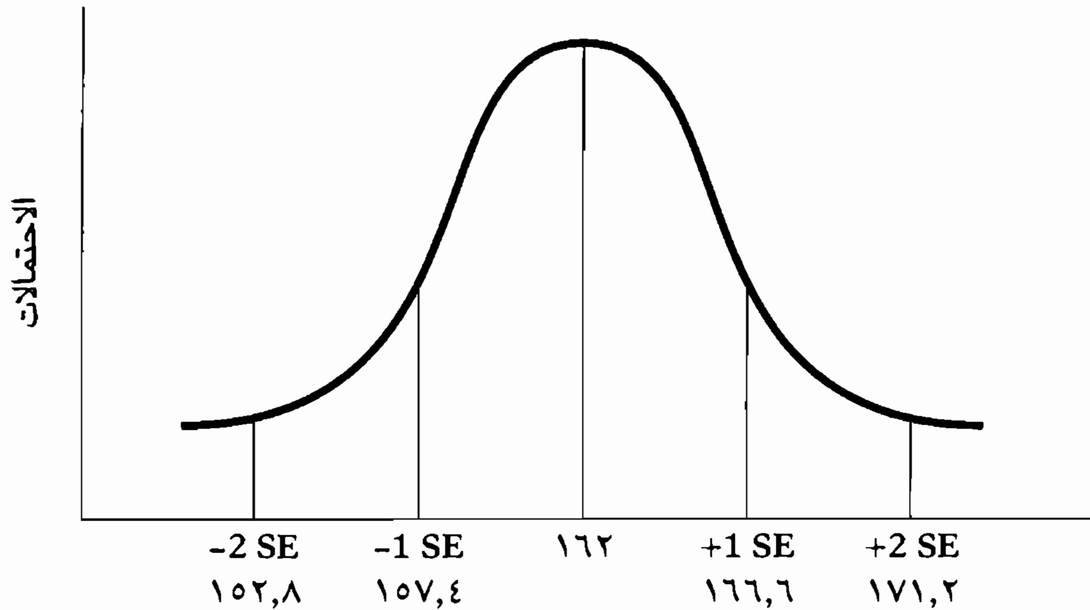
دعنا نتأمل مستوى الدلالة المقدر بـ ٠٫٠٥. يمكننا أن نرفض الفرضية المنعدمة عند مستوى ٠٫٥٠. أي إن هناك فرصة نسبتها ٥ بالمائة لأن نخرج بناتج يتساوى على الأقل في تطرفه مع ما كنا سنلحظه لو صحت الفرضية. يمكن لمثال بسيط أن يجعل الأمر أكثر وضوحًا. لكَمْ أكره أن أفعل ذلك بك مرة أخرى، لكن لنفترض أنك ذهبت مرة أخرى في مهمة بحث عن حافلة مفقودة (ربما بسبب جهودك الباسلة التي بذلتها في الفصل السابق). لكنك هذه المرة صرت تعمل بدوام كامل لدى باحثي دراسة الحيوات المتغيرة، وقد منحوك بيانات ممتازة تساعدك في عملك. وفي كل حافلة ركَّب الباحثون نحو ٦٠ راكبًا، وهو ما يعني أننا نستطيع اعتبار ركاب كل حافلة من الحافلات عينة عشوائية اتُّخذت من كامل مجتمع دراسة الحيوات المتغيرة. وذات صباح أفقتَ من نومك على خبر اختطاف إحدى الحافلات في محيط منطقة بوسطن من قِبل جماعة إرهابية مناصرة للبدانة*. مهمتك هنا تتلخص في أنك سوف تسقط من طائرة مروحية إلى سقف تلك الحافلة وهي تتحرك، ثم تتسلل إلى داخلها عبر مخرج الطوارئ، ثم تقرر بينما أنت متخفٍّ إذا ما كان راكبوها هم

* هذا المثال مستوحًى من أحداث حقيقية. لكن قمنا بالطبع بتغيير كثير من التفاصيل لأسباب تتعلق بالأمن القومي. وإنني لا أستطيع أن أؤكد أو أنفي اشتراكي في هذا الحدث.

المشاركين في دراسة الحيوات المتغيرة أم لا بناءً فقط على أوزانهم. (صدقني لا تختلف تلك الحبكة كثيرًا في عدم منطقيتها عن حبكات معظم أفلام الحركة والمغامرة، لكنها تتميز عنهم في كونها أكثر تثقيفًا).

وقت أن أقلعت الطائرة المروحية من قاعدة قوات الكوماندوز، تم تزويدك بمدفع رشاش وعدد من القنابل اليدوية وساعة يد تعمل كذلك ككاميرا ذات مستوى تصوير عالي النقاء، وكذلك بالبيانات التي حسبناها في الفصل السابق عن متوسط الوزن والانحراف المعياري للعينات التي انتقيت من بين المشاركين في دراسة الحيوات المتغيرة. فأي عينة عشوائية قوامها ٦٠ مشاركًا سيكون لها متوسط وزن متوقع يبلغ ١٦٢ رطلًا، وانحراف معياري يبلغ ٣٦ رطلًا، بما أن هذين هما المتوسط والانحراف المعياري لجميع المشاركين في الدراسة (كامل مجتمعها). من هذه البيانات نستطيع أن نحسب الانحراف المعياري لمتوسط العينة بهذا الشكل: $s/\sqrt{n} = 36/\sqrt{60} = 36/7.75 = 4.6$ في مقر العمليات تم نسخ التوزيع التالي داخل شبكية عينك اليمنى، حتى يتسنى لك أن ترجع إليه بعد أن تخترق الحافلة المتحركة وتقوم خفية بوزن جميع من بداخلها من ركاب.

توزيع متوسطات العينة



كما يظهر التوزيع السابق، نتوقع أن يكون لنحو ٩٥ بالمائة من العينة المكونة من ٦٠ شخصًا، والمنتقاة من دراسة الحيوات المتغيرة، متوسطُ وزنٍ يقع في نطاق انحرافين معياريين لمتوسط كامل مجتمع الدراسة، أو تقريبًا ما بين ١٥٣ رطلًا و١٧١ رطلًا*.

* لمزيد من التحديد نقول إن ٩٥ بالمائة من جميع متوسطات العينات ستقع في نطاق ١,٩٦ انحراف معياري فوق متوسط مجتمع الدراسة أو تحته.

وبالعكس فإنه في ٥ حالات فقط من أصل ١٠٠ حالة سيكون لعينة قوامها ٦٠ شخصًا، مأخوذة من المشاركين في دراسة الحيوات المتغيرة، متوسط أوزان أكبر من ١٧١ رطلًا أو أقل من ١٥٣ رطلًا. (إنك تُجري هنا ما يسمى باختبار الفرضية «ذي الذيلين»، والفارق بينه وبين الاختبار «ذي الذيل الواحد» سنتناوله في ملحق في نهاية هذا الفصل). قرر قادة القوة التي تشرف عليك في هذه العملية أن مستوى الدلالة في هذه المهمة هو ٠٫٠٥. إذا كان متوسط أوزان الركاب الستين في الحافلة المخطوفة يزيد عن ١٧١ أو ينقص عن ١٥٣، حينها سترفض الفرضية المنعدمة التي تقول بأنها تحمل المشاركين في الدراسة، وتقبل الفرضية البديلة التي تقول بأنها تحمل ستين راكبًا متجهين إلى مكان آخر، وستتمهل انتظارًا لأوامر جديدة.

نجحتَ في الهبوط فوق الحافلة المتحركة ووزنت الركاب خفية. ثم وجدت أن متوسط أوزان تلك العينة المكونة من ٦٠ راكبًا يبلغ ١٣٦ رطلًا؛ مما يعني أنه يقع تحت المتوسط بأكثر من انحرافين معياريين. (وكان هناك دليل آخر وهو أن جميع الركاب كانوا أطفالًا يرتدون قمصانًا رياضية مكتوبًا عليها «معسكر جلينديل للهوكي»).

طبقًا للتعليمات التي تلقيتها، عليك أن ترفض الفرضية المنعدمة التي تنص على أن هذه الحافلة التي تحوي عينة قوامها ٦٠ مشاركًا في دراسة الحيوات المتغيرة إذا ما وجدت مستوى الدلالة قد بلغ ٠٫٠٥. وهو ما يعني ما يلي: (١) أن متوسط الأوزان في تلك الحافلة يقع في نطاق لا نتوقع أن نشهده إلا في ٥ حالات من أصل ١٠٠، إذا صحت الفرضية المنعدمة وكانت الحافلة بالفعل تحمل المشاركين في الدراسة. (٢) تستطيع أن ترفض الفرضية المنعدمة عندما يبلغ مستوى الدلالة ٠٫٠٥. (٣) فيما متوسطه ٩٥ مرة من أصل ١٠٠، سيكون رفضك للفرضية المنعدمة إجراءً صحيحًا، وفي حوالي ٥ مرات من أصل ١٠٠ ستكون مخطئًا، بمعنى أنك قد اعتقدت أن تلك الحافلة ليست هي حافلة المشاركين في الدراسة في حين أنها كانت الحافلة المنشودة. وقد تصادف أن متوسط أوزان عينة المشاركين في الدراسة مرتفع أو منخفض نسبةً إلى متوسط أوزان جميع المشاركين في الدراسة.

لم تنتهِ المهمة بعد. فقد طلبتْ منك المسئولة عنك في المهمة (والتي تلعب أنجلينا جولي دورها في الفيلم المستوحى من ذلك المثال) أن تحسب قيمة الاحتمالات (p- value) للنتائج التي حصلتَ عليها. وقيمة الاحتمالات هي الاحتمالات المحددة للخروج بنتائج تُماثِل، على أقل تقدير، تطرُّفَ نظيرتها التي كنت ستلاحظها إذا ما صحت الفرضية المنعدمة. متوسط أوزان الركاب على هذه الحافلة يبلغ ١٣٦ رطلًا أو ٥٫٧ انحرافات معياريَّة أسفل متوسط أوزان المشاركين في الدراسة. واحتمالات الحصول على نتيجة بذلك التطرف، على الأقل إذا كانت تلك بالفعل حافلة المشاركين، هي أقل من ٠٫٠٠٠١. (في هذه الورقة البحثية ستكتب

النسبة بهذه الصيغة p<.0001). وبعد أن أنجزت مهمتك سوف تقفز من الحافلة فتستقر في المقعد الأمامي المجاور للسائق في سيارة مكشوفة تسير في الحارة المرورية المجاورة للحافلة.

[ولهذه القصة أيضًا نهاية سعيدة، فبعد أن علم أولئك الإرهابيون المنافحون عن البدانة عن وجود مهرجان للسجق في المدينة، اتفقوا على أن يهجروا العنف وأن يعملوا بشكل سلمي على الترويج للسمنة من خلال الترويج لمهرجانات السجق والتوسع فيها في جميع أرجاء العالم].

لو بدا لك مستوى الدلالة ٠٥. مستوًى اعتباطيًّا نوعًا ما، فذلك لأنه اعتباطي بالفعل. فليس هناك من حد موحد واحد ترفض عنده الفرضيات. والحدان ٠١. و١ كلاهما من الحدود الشائعة لمثل ذلك التحليل الذي ذكرناه لتوِّنا.

ومن الواضح أن رفض الفرضية المنعدمة عند مستوى ٠١. (يعني أن هناك فرصة تقل عن ١ من كل ١٠٠ أن تلاحظ نتيجة في ذلك النطاق إذا ما صحت الفرضية المنعدمة) له وزن إحصائي أكبر من رفض الفرضية المنعدمة عند مستوى ١ (بمعنى أن هناك فرصة من أصل عشر فرص لملاحظة هذه النتيجة إذا صحت الفرضية المنعدمة). سوف نناقش محاسن ومساوئ مستويات الدلالة المختلفة لاحقًا في هذا الفصل. أما النقطة المهمة الآن هي أننا عندما نتمكن من أن نرفض الفرضية المنعدمة عند مستوًى معين من الدلالة، حينها توصف النتائج بأنها «ذات دلالة إحصائية».

لكن هذا هو معناها في الحياة الواقعية. عندما تقرأ في الصحيفة أن من يتناولون عشرين كعكة من كعكات النخالة يوميًّا تنخفض معدلات الإصابة بسرطان القولون لديهم عنها لدى من لا يلتهمون ذاك الكم الهائل من النخالة، فلا بد أن البحث العلمي الذي استندت له الصحيفة قد بدا على هذا النحو: (١) من خلال مجموعة ضخمة من البيانات قرر الباحثون أن الأشخاص الذين يأكلون عشرين كعكة نخالة على الأقل في اليوم الواحد تنخفض لديهم حالات الإصابة بسرطان القولون عمن لم يعرف عنهم التهام كل تلك النخالة. (٢) نصت فرضية الباحثين المنعدمة على أن تناول النخالة لا علاقة له بسرطان القولون. (٣) ذلك التفاوت في النتائج المتعلقة بسرطان القولون، بين من يتناولون الكثير من النخالة ومن لا يفعلون، لا يمكن أن يفسر بأنه مجرد صدفة. وبشكل أكثر تحديدًا إذا لم يكن هناك أي ارتباط حقيقي بين تناول كعك النخالة وسرطان القولون، فإن احتمالات حدوث تلك الفجوة الواسعة بين ملتهمي النخالة وغيرهم عن طريق الصدفة فقط لا يرقى لأن يصل إلى حد معين، كالحد ٠٥. مثلًا. (يجب على الباحثين أن يعينوا ذلك الحد قبل أن يجروا تحليلهم الإحصائي كي يتفادوا اختيار الحد الأفضل في جعل نتائجهم تبدو أكثر دلالة). (٤) ولا بد أن البحث الأكاديمي سيتضمن خلاصة تشابه ما في السطور التالية:

«لقد وجدنا ارتباطًا ذا دلالة إحصائية بين استهلاك عشرين كعكة نخالة أو أكثر يوميًّا وانخفاض تواتر حالات الإصابة بسرطان القولون. وهذه النتائج تصير ذات دلالة عند مستوى ٠٫٠٥».

عندما سأقرأ عن تلك الدراسة في وقت لاحق في صحيفة شيكاجو صن تايمز، بينما أتناول فطوري المكون من لحم الخنزير المقدد والبيض، غالبًا سأجد ذلك العنوان الصحفي مباشرًا ومثيرًا للاهتمام: «عشرون كعكة نخالة في اليوم ستبعد عنك سرطان القولون». لكن رغم أن ذلك العنوان الصحفي سيكون أكثر جذبًا للاهتمام من الورقة البحثية، فإنه قد ينطوي كذلك على مستوًى خطير من عدم الدقة. فالدراسة من الناحية الواقعية لا تدَّعي أن تناول كعكات النخالة يجعل الإنسان أقل عرضة للإصابة بسرطان القولون، وإنما هي لا تفعل أكثر من إظهار الارتباط السلبي بين استهلاك كعكات النخالة وتواتر حدوث حالات سرطان القولون الذي يظهر في مجموعة كبيرة من البيانات. لكن الارتباط الإحصائي غير كافٍ لإثبات أن كعكات النخالة تُسبِّب تحسنًا في الناتج الصحي. فمن الممكن أن يكون أولئك الذين يأكلون كعكات النخالة (وخاصة من يأكلون عشرين منها يوميًّا) يفعلون أمورًا أخرى تقلل من فرص إصابتهم بالسرطان؛ كأن يقللوا من استهلاك اللحوم الحمراء، ويمارسوا الرياضة بانتظام، وإجراء الفحوصات الخاصة بذلك المرض، وما إلى ذلك. (هذا هو «انحياز المستخدم الصحي» الذي جاء في الفصل السابع). هل يكمن السبب في كعكات النخالة، أم هي سلوكيات أخرى وخصال شخصية تَصادَف أن من يلتهمون تلك الكميات من كعك النخالة يَتبنَّوْنها؟ ذلك الفارق بين الارتباط والسببية هو أمر جوهري في الوصول إلى التفسير الصحيح للنتائج الإحصائية. سوف نعود لفكرة أن «الارتباط لا يكافئ السببية» في مكان لاحق في الكتاب.

كما يجب عليَّ أن أشير هنا إلى أن الدلالة الإحصائية ليس لها علاقة بحجم الارتباط. فمن يأكلون الكثير من كعكات النخالة قد تنخفض معدلات إصابتهم بسرطان القولون، لكن كم قدر ذلك الانخفاض؟ فقد يكون الفارق في معدلات الإصابة بالسرطان بين آكلي النخالة ومَن لا يأكلونها فارقًا لا يُذكر، ونتيجة الدلالة الإحصائية تعني فقط أن ذلك التأثير الملاحظ، أيًّا كانت ضآلته، لا يرجَّح حدوثه عن طريق الصدفة. لنفترض أنك عثرت على دراسة محكمة التصميم قد وَجدت علاقة إيجابية ذات دلالة إحصائية بين تناول موزة قبل اختبار SAT وتحقيق علامات جيدة في قسم الرياضيات من الاختبار. هنا قد ترغب في السؤال عن عدة أمور، وواحد منها هو كم يبلغ قدر ذلك التأثير؟ من الممكن جدًّا أن يبلغ قدره ٩ نقاط في اختبار متوسط إجاباته ٥٠٠، وهو الرقم الذي لن يحوِّل مسار حياة الطالب. في الفصل الحادي عشر سوف نعود إلى ذلك الفارق الجوهري بين الحجم والدلالة عندما يتعلق الأمر بنتائج إحصائية.

لكن في نفس الوقت قد يكون الوصول إلى نتيجة مفادها أنه «لا يوجد ارتباط ذو دلالة إحصائية» بين اثنين من المتغيرات، يعني أن أي علاقة بين المتغيرين يمكن أن تفسَّر بشكل منطقي عن طريق الصدفة وحدها. قامت صحيفة النيويورك تايمز مؤخرًا بنشر تحقيق فاضح لشركات البرامج التكنولوجية التي تروِّج لبرامج من إنتاجها تدعي قدرتها على تحسين أداء التلاميذ، رغم كون البيانات تشير إلى أمر آخر[٣] يذكر المقال أن جامعة كارنيجي ميلون تبيع برنامجًا يسمى المعلم الإدراكي، مدعية ادعاءً جريئًا بأنه: «ثورة في مناهج الرياضيات، وثورة في نتائجها». لكن تقييم ذلك البرنامج الذي قامت به وزارة التعليم الأمريكية قد خلص إلى أنه ليست له «أي آثار واضحة» على نتائج اختبارات تلاميذ المدارس الثانوية. (اقترحت صحيفة التايمز تغيير شعار حملة الترويج للبرنامج ليكون «مناهج رياضية لا يميزها شيء، ونتائج غير موثوق بها»). بل إنه قد تم تقييم عشرة برامج قد صممت لتعليم مهارات في مجالات الرياضيات والقراءة، فوُجد أن تسعة منها «ليس لها آثار ذات دلالة إحصائية على نتائج الاختبارات». بعبارة أخرى نقول إن الباحثين الفدراليين لا يستطيعون استبعاد الصدفة المحضة كمسبب للتغاير الذي يحدث بين أداء التلاميذ الذين يستخدمون تلك البرامج وزملائهم الذين لا يستخدمونها.

سأتوقف هنا لأذكِّرك بأهمية كل ما سبق. فقد نُشر مقال بصحيفة وول ستريت جورنال في شهر مايو من عام ٢٠١١ بعنوان: «الرابط بين التوحد وحجم المخ». لقد كان ذلك فتحًا عظيمًا، فلطالما ظلت أسباب اضطراب طيف التوحد غامضة. جاءت أولى سطور ذلك المقال ملخصًا لدراسة نشرت في دورية Archives of General Psychiatry قائلة: «إن الأطفال المصابين بالتوحد لهم أمخاخ أكبر حجمًا من أمخاخ الأطفال الذين لا يعانون ذلك الاضطراب، وقد أفادت دراسة حديثة صدرت الإثنين الماضي بأن نمو الحجم ذلك يحدث قبل عمر السنتين[٤]». فقد أثبتت الصور الدماغية التي أجريت على أمخاخ ٥٩ طفلًا مصابين بالتوحد و٣٨ غيرهم لم يصابوا بالمرض، وقرر باحثو جامعة نورث كارولينا أن الأطفال المتوحدين لهم أمخاخ تزيد في حجمها بنسبة ١٠ بالمائة عن نظائرها لدى الأطفال غير ذوي التوحد.

إليك السؤال الطبي المناسب هنا: أهناك فارق فسيولوجي يميز أمخاخ الأطفال المصابين باضطراب طيف التوحد؟ لو كان ثمة فارق بالفعل، لأمكن أن يقودنا هذا إلى فهمٍ أفضل لأسباب ذلك الاضطراب وكيفية علاجه أو الوقاية منه.

وإليك السؤال الإحصائي المناسب هنا: هل يستطيع الباحثون أن يقوموا بتلك الاستدلالات المذهلة المتعلقة باضطراب طيف التوحد اعتمادًا بشكل عام على دراسة ما قد تبدو مجموعة صغيرة من الأطفال المصابين بالمرض (٥٩) ومجموعة ضابطة أقل عددًا (٣٨) – أو ٩٧ طفلًا إجمالًا؟ الإجابة هي نعم. استخلص الباحثون أن احتمالات ملاحظة

الفوارق التي وجدوها في إجمالي عدد أحجام الأمخاخ في العينتين سيكون فقط ٢ في ١٠٠٠ ($p = .002$)، إذا لم يكن هناك بالفعل فارق حقيقي بين أحجام أمخاخ جميع الأطفال المتوحدين وغيرهم من غير المتوحدين جميعًا.

قمت بالبحث عن الدراسة الأصلية التي نشرت في الدورية العلمية التي ذكرناها سابقًا[٥]. فوجدت أن المنهجيات التي استخدمها أولئك الباحثون ليست أكثر تعقيدًا من المفاهيم التي تناولناها حتى الآن. سوف أصطحبكم في جولة سريعة أعرِّفكم فيها بالأساسات التي قامت عليها هذه النتائج المهمة من الناحيتين الاجتماعية والإحصائية. في البداية لا بد أن تأخذ في اعتبارك أن كل مجموعة من هاتين المجموعتين، سواء مجموعة الأطفال الـ ٥٩ المتوحدين أومجموعة الأطفال الـ ٣٨ غير المتوحدين، تتكون من عينة كبيرة نسبيًّا منتقاة بشكل صحيح من المجتمع الذي تمثله، والذي هو كل الأطفال المصابين وغير المصابين باضطراب طيف التوحد. وهذه العينات من الكبر أن تنطبق عليها النهاية المركزية. وإذا ما كنت قد حاولت أن تحذف الفصل السابق من ذهنك، فسوف أذكِّرك أنا بما وضَّحتْه لنا مبرهنة النهاية المركزية: (١) متوسطات العينات المنتقاة من أي مجتمع كان سوف تتوزع بشكل تقريبي على أنها توزيع طبيعي حول متوسط كامل المجتمع. (٢) من المتوقع أن يكون متوسط العينة وانحرافها المعياري مساويًا تقريبًا لمتوسط كامل المجتمع الذي انتقيت منه وانحرافه المعياري، (٣) نحو ٦٨ بالمائة من متوسطات العينات ستقع في نطاقِ خطأ معياري واحد من متوسط المجتمع، وسيقع نحو ٩٥ بالمائة في نطاق خطأين معياريين من متوسط المجتمع، وهلم جرًّا.

لتبسيط الأمر بلُغة أقل تقنية نقول إن ما سبق يعني أن أية عينة يجب أن تشابه المجتمع الذي انتقيت منه، ورغم أن كل عينة تختلف عن الأخرى، فإنه لَأمر نادر نسبيًّا أن ينحرف متوسط عينة منتقاة جيدًا انحرافًا كبيرًا عن متوسط كامل المجتمع الأساسي. وبالمثل فإننا نتوقع أن تتشابه أي عينتين انتُقيَتا من مجتمع واحد إحداهما مع الأخرى. أو من الممكن أن ننظر إلى الأمر بطريقة أخرى فنقول إننا إن وجدنا عينتين متباينتين أشد التباين في متوسطيهما، فحينها يكون التفسير الأرجح لهذا هو أنهما منتقتان من مجتمعين مختلفين.

إليك هذا المثال المنطقي السريع. لنفترض أن فرضيتك المنعدمة تنص على أن متوسط طول لاعبي كرة السلة المحترفين يماثل نظيره لدى كامل مجتمع الذكور البالغين. تقوم حينها بانتقاء عشوائي لعَينة من خمسين لاعبًا لكرة السلة محترفًا وعينة أخرى من خمسين ذكرًا بالغًا لا يحترفون تلك اللعبة. لنفترض أن متوسط الطول في عينة لاعبي البيسبول يبلغ ٦ أقدام و٧ بوصات، ومتوسط أطوال المجموعة الأخرى يبلغ ٥ أقدام و١٠ بوصات (ما يعني فارقًا يبلغ ٩ بوصات). ما احتمالات ملاحظة فارق شاسع بين متوسط

أطوال المجموعتين، إذا لم يكن هناك فارق واقعي بين متوسطات أطوال لاعبي كرة السلة المحترفين وغيرهم في كامل المجتمع؟ الإجابة غير التقنية هنا هي: احتمالات ضعيفة جدًّا جدًّا جدًّا*.

نفس تلك المنهجية الأولية انتهجتها دراسة التوحد. فقد قامت الدراسة بالمقارنة بين مقاييس مختلفة لأحجام الأمخاخ في عينات من الأطفال. (تمت قياسات الأمخاخ تلك عن طريق صور الرنين المغناطيسي التي أُخذت للأطفال في عمر السنتين، ثم أخذت مجددًا عندما كانوا بين الرابعة والخامسة). سوف أركز على مقياس واحد فقط، ألا وهو حجم المخ الكلي. نصت الفرضية المنعدمة التي وضعها الباحثون على عدم وجود أي فارق تشريحي بين أحجام أمخاخ الأطفال ذوي التوحد وغيرهم. أما الفرضية البديلة فقد كانت تقضي بوجود اختلاف أساسي في أمخاخ الأطفال المصابين باضطراب طيف التوحد. صحيح أن تلك النتيجة لم تُجب عن الكثير من الأسئلة، إلا أنها أضاءت الطريق إلى اتجاه للبحث المتقدم.

في تلك الدراسة وجد أن الأطفال ذوي اضطراب طيف التوحد يبلغ متوسط حجم أمخاخهم ١٣١٠,٤ سنتيمترات مكعبة، أما أطفال المجموعة الضابطة فقد كان متوسط أحجام أمخاخهم ١٢٣٨,٨ سنتيمترًا مكعبًا. وهكذا فإن الفارق بين متوسطَي أحجام الأمخاخ في المجموعتين يبلغ ٧١,٦ سنتيمترًا مكعبًا. إلى أي مدى من المرجح أن نخرج بهذه النتيجة إذا لم يكن هناك فارق عام بين كامل مجتمع الأطفال الذين يعانون من اضطراب طيف التوحد، وبين كامل مجتمع الأطفال الذين لا يعانونه؟

لعلك تذكر من الفصل السابق أننا نستطيع إيجاد الخطأ المعياري لأي عينة من عيناتنا بهذه الصيغة: $s/\sqrt{n}$ ؛ حيث s هو الانحراف المعياري للعينة، وn هو عدد المشاهدات. قدمت لنا تلك الدراسة كل تلك الأرقام. فالخطأ المعياري لأحجام أمخاخ كامل عينة الأطفال ذوي التوحد التي قوامها ٥٩ طفلًا قدره ١٣ سنتيمترًا مكعبًا، أما الخطأ المعياري لإجمالي عينة الضبط المكونة من ٣٨ طفلًا، فقدره ١٨ سنتيمترًا مكعبًا. ولعلك تذكر أن مبرهنة النهاية المركزية ستنبئنا بأنه في ٩٥ عينة من أصل ١٠٠ سوف سيقع متوسط العينة في نطاق خطأين معياريين من المتوسط الفعلي لكامل المجتمع في أي من الاتجاهين.

لدينا الآن عينة وحيدة ذات متوسط يبلغ ١٣١٠,٤ سنتيمترات مكعبة. فإذا ما

* هناك فرضيتان بديلتان ممكنتان. واحدة منهما تنص على أن لاعبي كرة السلة الذكور يفوقون بقية مجتمع الذكور طولًا. أما الثانية فتقضي بأن لاعبي كرة السلة المحترفين لهم متوسط أطوال مختلف عن نظيره لدى بقية مجتمع الذكور (وهي التي تفتح إمكانية كون لاعبي كرة السلة الذكور أقصر طولًا من بقية الرجال). لسوف نجد لهذا الفارق تأثيرًا صغيرًا عندما نجري اختبارات الدلالة ونحسب قيمة الاحتمالات. هذا الأمر مشروح في كتب أخرى متقدمة لكنه غير ذي أهمية في مناقشتنا العامة تلك.

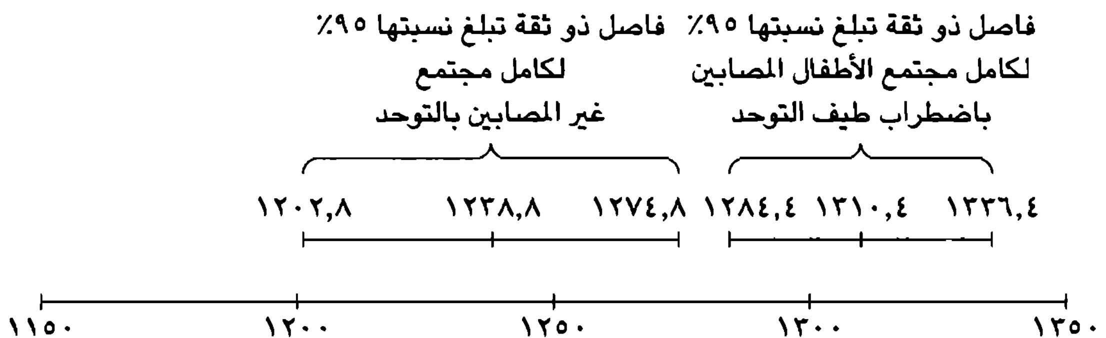
وضعنا فاصلًا يمتد لخطأين معياريين في كلا الاتجاهين (١٣١٠,٤ ± ٢٦) فسَيَتسنَّى لنا أن نتيقن استنادًا للمنطق من أن ذلك الفاصل سيتضمن متوسط أحجام أمخاخ جميع الأطفال ذوي اضطراب طيف التوحد. إلى مدى سيكون ذلك اليقين؟

بل إن مبرهنة النهاية المركزية تقول لنا إننا لو واصلنا انتقاء العينة تلو الأخرى من المجتمع كاملًا، فسوف يحتوي نحو ٩٥ بالمائة من الفواصل التي حسبناها على المدى الطويل على المتوسط الحقيقي لكامل المجتمع. وبهذا تفترض الاحتمالية أن الفاصل الذي حسبناه من خلال عينتنا هو واحد من الـ ٩٥ بالمائة «الجيدة»، لا الـ ٥ بالمائة «السيئة».

هذا التمرين بأكمله يخلق ما يسمى بفاصل الثقة. فإننا نستطيع أن نقول بثقة قدرها ٩٥ بالمائة إن النطاق الذي يتدرج من ١٢٨٤,٤ إلى ١٣٣٦,٤ سنتيمترًا مكعبًا (متوسط العينة البالغ ١٣١٠,٤ ± خطأين معياريين)، يحتوي على متوسط أحجام أمخاخ كامل مجتمع الأطفال المصابين باضطراب طيف التوحد.

باستخدام نفس المنهجية نستطيع أن نقول بنسبة ثقة قدرها ٩٥ بالمائة إن الفاصل البالغ ١٢٣٨,٨ ± ٣٦ أو ما بين ١٢٠٢,٨ و١٢٧٤,٨ سنتيمترًا مكعبًا، سوف يحتوي على متوسط حجم أمخاخ الأطفال في كامل مجتمع الأطفال الذين لا يعانون اضطراب طيف التوحد.

نعم تلك أرقام كثيرة، بل ولعلك قد قذفت الكتاب بطول يدك إلى نهاية الغرفة*. لكن إن لم تكن قد فعلت ذلك أو فعلته ثم ذهبت والتقطت الكتاب مجددًا، فعليك أن تلاحظ هنا أن فواصل الثقة التي لدينا هنا لا تتداخل. فالحد الأدنى لفاصلنا ذي الثقة البالغة نسبتها ٩٥ بالمائة لمتوسط أحجام أمخاخ كامل مجتمع الأطفال المصابين بالتوحد (١٢٨٤,٤ سنتيمترًا مكعبًا) لا يزال أعلى من الحد الأعلى لفاصل الثقة البالغة نسبتها ٩٥ بالمائة لمتوسط أحجام أمخاخ كامل مجتمع الأطفال غير المصابين بالتوحد (١٢٧٤,٨ سنتيمترًا مكعبًا) كما هو موضح بالشكل التالي.



* أعترف أنني ذات مرة مزقت أحد كتب الإحصاء فشطرته إلى نصفين.

ذلك هو الدليل الأول على أنه قد يكون هناك اختلاف تشريحي في أمخاخ الأطفال المصابين باضطراب طيف التوحد. لكنه ما زال مجرد دليل، فكل هذه الاستدلالات قائمة على بيانات أقل من ١٠٠ طفل، وربما كانت عيناتنا كلها غير معتادة.

يتبقى إجراء إحصائي أخير قادر على أن يؤتي ثماره. لو كان الإحصاء حدثًا أولمبيًّا، مثل التزلج الفني على الجليد، لصار ذلك الإجراء هو البرنامج الأخير الذي ترمي فيه الجماهير المبتهجة بصُحبات الورود على الجليد. نستطيع أن نحسب بالضبط احتمالات ملاحظة الاختلاف في المتوسطات، على الأقل ذلك الاختلاف الكبير (١٣١٠,٤ سنتيمترات مكعبة في مقابل ١٢٣٨,٨ سنتيمترًا مكعبًا)، إذا لم يكن هناك بالفعل اختلاف في أحجام أمخاخ الأطفال المصابين بالتوحد وغيرهم على مستوى المجتمعين كاملين. إننا قادرون على أن نحسب قيمة الاحتمالات للاختلاف الملاحظ في المتوسطات.

ولأنني أخشى أن تقذف الكتاب إلى ركن الغرفة مرة أخرى، فقد وضعت المعادلة في ملحق هذا الفصل. المنطق في هذا مستقيم جدًّا. إذا ما انتقينا عينتين كبيرتين من مجتمع واحد، فلا بد أن نتوقع تشابهًا كبيرًا في متوسطيهما. بل إن أفضل تخميناتنا سيفترض أن يتطابق المتوسطان. فمثلًا إذا ما انتقيت ١٠٠ لاعب من البطولة الوطنية لكرة السلة، وكان متوسط أطوال تلك المجموعة ٦ أقدام و٧ بوصات، فلي أن أتوقع أنني إذا انتقيت عينة عشوائية أخرى من نفس المجتمع قوامها ١٠٠ لاعب أيضًا، فسيكون متوسط الأطوال فيها أيضًا ٦ أقدام و٧ بوصات. قد يكون هناك فارق بين متوسطي العينتين بقدر بوصة أو اثنتين، لكنه من غير المرجح أن يزيد ذلك الفارق إلى أربع بوصات، ومن المستبعد جدًّا أن يتسع الفارق إلى ستة بوصات أو ثمانية. لقد اتضح أننا قادرون على أن نحسب الخطأ المعياري للفارق بين متوسطين لعَينتين، وهذا الخطأ المعياري يمنحنا مقياسًا للتشتت الذي نتوقعه بشكل تقريبي عندما نطرح أحد المتوسطين من الآخر. (المعادلة في ملحق الفصل كما ذكرنا آنفًا). لكن الأمر المهم هنا هو أننا نستطيع استخدام هذا الخطأ المعياري لحساب احتمالات كون هاتين العينتين من نفس المجتمع. فيما يلي أعرض لك الكيفية التي يسير ذلك الإجراء بها:

١- إذا ما كانت العينتان منتقتين من نفس المجتمع، فإن أفضل توقعاتنا يقدِّر الفارق بين متوسطيهما بصفر.

٢- تنبئنا مبرهنة النهاية المركزية بأنه في العينات المكررة يتوزع الفارق بين المتوسطين بشكل تقريبي كتوزيع طبيعي. (هل صرت تحب مبرهنة النهاية المركزية أم لم تفعل بعد؟)

٣- إذا ما كانت العينتان قد انتُقيتا بالفعل من مجتمع واحد، فسوف يكون الفارق بين العينيتن، في ٦٨ حالة من أصل ١٠٠، واقعًا في نطاق خطأ معياري واحد من الرقم

صفر. وفي ٩٥ حالة تقريبًا من أصل ١٠٠ حالة، يقع الفارق بين متوسطي العينتين في نطاق خطأين معياريين من الرقم صفر. وفي ٩٩٫٧ حالة من أصل ١٠٠ يقع الفارق في نطاق ثلاثة أخطاء معيارية من الرقم صفر – وهو الأمر الذي شجع على الخروج بنتيجة دراسة التوحد تلك التي بدأنا بها.

كما ذكرنا من قبل، يبلغ الفارق في متوسط عينتَي أحجام أمخاخ الأطفال ذوي التوحد وغيرهم ممن تحتويهم عينة التحكم ٧١٫٦ سنتيمترًا مكعبًا. ويبلغ الخطأ المعياري لهذا الفارق ٢٢٫٧، مما يعني أن الفارق في المتوسطات بين العينيتين يتعدى ثلاثة أخطاء معيارية من الرقم صفر، ولنا أن نتوقع ناتجًا بهذا التطرف (أو أكثر منه تطرفًا) في حالتين فقط من أصل ١٠٠٠ حالة، إذا كانت العينتان منتقتين من ذات المجتمع.

قرر باحثو الدراسة التي نشرت في دورية Archives of General Psychiatry أن قيمة الاحتمالات تبلغ ٠٫٠٠٢. كما ذكرت آنفًا. الآن تعرف أنت من أين أتت تلك القيمة.

لكن مع كل العجائب التي يأتي بها الاستدلال الإحصائي، فإنه له بعض المثالب الجسيمة. وهذه المثالب نستطيع أن نراها من خلال ذاك المثال الذي استهللنا به هذا الفصل حول أستاذي المستريب. فعملية الاستدلال الإحصائي القوية تلك تقوم على الاحتمالات، لا على أي نوع من أنواع اليقين العلمي. إننا لا نرغب في الزج بالناس في السجون لمجرد حدوث أمر هو أقرب لسحب ورقتين رابحتين على التوالي في لعبة ورق، فمن الممكن أن يحدث ذلك حتى لو لم يكن ذلك ينطوي على غش. ونتيجة لهذا تتبدى أمامنا معضلة أساسية عندما نصل لمرحلة اختبار أي فرضية.

تبدت هذه الحقيقة الإحصائية عام ٢٠١١ عندما تجهزت دورية Journal of Personality and Social Psychology لإصدار ورقة بحثية بدت ظاهريًّا كأنها آلاف الأوراق البحثية لا واحدة[٦]. قام أحد أساتذة جامعة كورنيل بالإعلان عن فرضية منعدمة، وقام بإجراء تجربة لاختبار تلك الفرضية المنعدمة، ثم رفض تلك الفرضية المنعدمة عند مستوى دلالة قدره ٠٫٠٥. استنادًا إلى نتائج تلك التجربة. أثارت تلك النتائج ضجة كبيرة في الدوائر العلمية، وكذلك في وسائل الإعلام كصحيفة النيويورك تايمز.

يكفي أن نقول إن مقالات تلك الدورية لا تجتذب دائمًا عناوين الأخبار العريضة. فما الذي جعل هذه الدراسة مثيرة للجدل إلى هذا الحد؟ كان السبب هو أن ذاك الباحث الذي نقصده يختبر قدرة البشر على تمرير الإدراك الفائق الحس الذي يختصر بحروف ESP. تنص الفرضية المنعدمة على عدم وجود ذلك النوع الفائق من الإدراك، أما الفرضية البديلة فتنص على أن البشر يتمتعون بقدرات حسية فائقة. لدراسة هذا السؤال جنَّد الباحث عينة كبيرة من المشاركين لدراسة «ستارَيْن» يغطيان شاشة جهاز كمبيوتر. وعلى هذه الشاشة برنامج يعرض بشكل عشوائي صورًا مثيرة جنسيًّا خلف واحد من الستارين. من

خلال المحاولات المتكررة استطاع المشاركون أن ينتقوا الستار الذي توارت خلفه الصورة الإباحية بنسبة ٥٣ بالمائة من المرات، في حين تقول الاحتمالات أنه يفترض بتلك النسبة أن تكون ٥٠ بالمائة. وبسبب كبر حجم العينة تمكَّن الباحث من رفض الفرضية المنعدمة التي تنص على عدم وجود الإدراك الفائق الحس، وقَبِل مكانها الفرضية البديلة التي تنص على أن الإدراك الحسي الفائق قادر على أن يجعل بعض الأشخاص يتنبئون بالأحداث المستقبلية. لكن قرار نشر تلك الدراسة قد لاقى انتقادًا واسع النطاق، استنادًا إلى أن أي حدث مفرد ذي دلالة إحصائية يمكن بكل سهولة أن يفسَّر على أنه نتاج لصدفة محضة، خاصة عندما لا يكون هناك أي دليل يعززه أو حتى يفسر تلك النتائج. تلخص صحيفة النيويورك تايمز تلك الانتقادات بقولها: «إن الادعاءات التي تتحدى جميع قوانين العلم تقريبًا تعتبر من خلال تعريفها استثنائية، ومن ثم فإنها تحتاج لدليل استثنائي. وتجاهُل ذلك الاعتبار – وهو ما يفعله التحليل القائم على علم الاجتماع – يجعل الكثير من النتائج تبدو ذات دلالة أكبر مما هي عليه في الحقيقة».

واحد من حلول التغلب على ذلك الضرب من الهراء يتمثل في وضع حد صارم لتحديد الدلالة الإحصائية كالحد ٠٫٠٠١. * لكن ذلك في حد ذاته يخلق مشكلة أخرى. فاختيار مستوًى مناسب من الدلالة يستلزم مساومة ضمنية.

إذا لم نكن نجد الكثير من الغضاضة في رفض الفرضية المنعدمة (كأن يكون الحد ٠٫١)، فسوف نجد أنفسنا نرفضها بين الحين والآخر، رغم أنها قد تكون صحيحة (أشك أن هذا هو الحال الذي جرى في دراسة الإدراك الحسي الفائق). وفي لغة الإحصاء يُعرف هذا باسم الخطأ من النوع ١. لنتأمل مثال قاعة المحكمة الأمريكية، حيث تنص الفرضية المنعدمة على أن المدعى عليه غير مذنب، ويكون حد رفض الفرضية المنعدمة هو «مذنب بما يتجاوز أي شك منطقي». لنفترض أننا ارتضينا بحد منخفض على غرار «لدينا شعور قوي بأن ذلك الرجل قد فعلها». فهذا سوف يضمن الزج بالمزيد من المجرمين إلى السجون – والمزيد من الأبرياء أيضًا. وإذا ما ترجمنا هذا الأمر في سياق إحصائي، فسنجده مكافئًا للتوقف عند مستوى دلالة متدنٍّ نسبيًّا كالمستوى ٠٫١

لكن الواقع أن احتمالًا يبلغ ١ من ١٠ هو ليس باحتمال مستبعد جدًّا عندما تضعه في سياق معين، مثل اعتماد عقَّار جديد لعلاج السرطان. فمن كل عشرة عقاقير نعتمدها من خلال عبء إثبات إحصائي متدنٍّ نسبيًّا، نجد واحدًا منها غير ذي جدوى فعلية، ونجده لم يُظهر تلك النتائج الواعدة إلا عن طريق الصدفة. (وإذا ما طبقنا هذا على مثال قاعة المحكمة، لوجدنا أنه من كل عشرة مدعى عليهم نراهم مذنبين، بهم واحد هو في الواقع بريء). فالخطأ من النوع ١ ينطوي على دحض نظرية منعدمة عن طريق الخطأ. وهو

* وهناك حل آخر يتمثل في محاولة استنساخ تلك النتائج في دراسات إضافية.

ما يطلق عليه أيضًا مصطلح «إيجابي زائف»، رغم كون هذا المصطلح منافيًا للمنطق. إليك إحدى الطرق التي يمكن أن نفهم من خلالها تلك الرطانة. عندما تذهب إلى الطبيب كي تخضع لفحص خشيةَ أن تكون مصابًا بمرض معين، حينها تنص الفرضية المنعدمة على أنك غير مصاب بالمرض. إذا ما أمكن استخدام النتائج المعملية لرفض الفرضية المنعدمة، فمعنى هذا أن فحوصك إيجابية. وإذا ما جاءت فحوصك إيجابية لكنك لم تكن بالفعل مصابًا بالمرض، ففحصك إيجابي زائف.

في أي حالة يكون فيها عبء إثبات الفرضية المنعدمة خفيفًا، تزيد أرجحية تحقُّقها. ومن الطبيعي ألا نفضِّل اعتماد عقاقير غير فعالة لعلاج السرطان أو إيداع مدعى عليهم أبرياء بالسجون.

لكن هنا يبرز التحدي، فكلما ارتفع حد رفض الفرضية المنعدمة، زادت أرجحية فشلنا في رفض تلك الفرضية المنعدمة التي يجب أن ترفض. فإذا ما تطلب الأمر خمسة شهود عيان لإدانة أي مدعًى عليه مجرم، فلسوف يطلق سراح الكثير من المذنبين عن باطل. (وبالطبع سيقل عدد الأبرياء الذين سيسجنون ظلمًا). وإذا ما تَبنَّيْنا مستوى الدلالة البالغ ٠٫٠٠١ في تجاربنا السريرية لعقاقير السرطان، فلسوف نحدُّ من اعتماد العقاقير غير الفعالة. (إن فرصة رفض الفرضية المنعدمة التي تقضي بأن العقَّار ليس بأكثر نجاعة من العقَّار الوهمي، خطأً، تقدر بـ ١ إلى ١٠٠٠). لكننا الآن نخاطر بعدم إقرار العديد من العقاقير الفعالة لأننا رفعنا سقف الاعتماد عاليًا جدًّا. وذلك ما يعرف بالخطأ من النوع ٢ أو السلبي الزائف.

أي الخطأين أفدح؟ يعتمد ذلك على الظروف المحيطة. لكن النقطة الأهم هنا هي أن تضع المساومة في اعتبارك. فليس هناك ما يعرف بالغداء المجاني الإحصائي. لنتأمل تلك المواقف غير الإحصائية، والتي تنطوي كلها على مساومة بين النوعين ١ و٢ من الخطأ.

١- مرشحات البريد الإلكتروني الضار. هنا تنص الفرضية المنعدمة على أن أي رسالة إلكترونية واردة لا تندرج تحت قائمة البريد الضار. يفتش مرشح الرسائل الإلكترونية الضارة عن أدلة في أي رسالة إلكترونية واردة قد تساعد على رفض الفرضية المنعدمة؛ كضخامة قائمة مَن أُرسلت إليهم تلك الرسالة، أو مثل عبارات واردة بها على غرار: «تضخيم العضو الذكري». سوف يؤدي الوقوع في خطأ من النوع الأول إلى استبعاد إحدى الرسائل الإلكترونية غير الضارة في الواقع (إيجابي زائف). أما الخطأ من النوع ٢ فسوف ينطوي على السماح بمرور رسالة ضارة من خلال المرشح إلى صندوق بريدك الإلكتروني (سلبي زائف). وبالنظر إلى كلفة عدم تلقِّي رسالة إلكترونية مهمة في مقابل كلفة تلقِّي رسالة عرضية بين الحين والآخر تروِّج للفيتامينات العشبية، فلا بد أن معظم الناس سيميلون إلى السماح

بالخطأ رقم ٢. لذلك فإن أي مرشح للبريد الضار مصمم بشكل مثالي، عليه أن يطلب مستوًى عاليًا من اليقين قبل أن يرفض الفرضية المنعدمة التي تقضي بعدم شرعية رسالة إلكترونية ما، ومن ثم صدها.

٢- فحوص مرض السرطان. هناك الكثير من الفحوصات التي تجرى لاكتشاف الأورام السرطانية؛ مثل فحوص سرطان الثدي، وفحوص سرطان البروستاتا، وحتى أشعة الرنين المغناطيسي على كامل الجسد، والتي تهدف إلى اكتشاف أي شيء يثير الريبة فيه. تنص الفرضية المنعدمة في حالة أيٍّ ممن يخضعون إلى تلك الفحوص على عدم وجود أي أورام سرطانية. تُستخدم هذه الفحوص لرفض النظرية المنعدمة في حالة وجود ما يريب في نتائجها. دائمًا ما يفترض الناس أن الوقوع في خطأ من النوع ١ (الإيجابي الزائف الذي يعني عدم وجود سرطان في الحقيقة)، هو أفضل بكثير من الوقوع في الخطأ من النوع ٢ (السلبي الزائف الذي يفشل في اكتشاف السرطان رغم وجوده). لو نظرنا للأمر تاريخيًّا لوجدنا مثال فحوص السرطان هو نقيض مثال مرشحات الرسائل الإلكترونية الضارة. فالأطباء والمرضى على السواء يفضلون السماح بعدد معقول من الأخطاء من النوع ١ (الإيجابي الزائف) في مقابل اجتناب احتمالات وقوع خطأ واحد من النوع ٢ (عدم اكتشاف ورم سرطاني موجود). لكن مؤخرًا بدأ خبراء السياسات الصحية في نفي هذه الرؤية بسبب التكاليف الباهظة والأعراض الجانبية الخطيرة التي تنطوي عليها النتائج الإيجابية الزائفة.

٣- الإيقاع بالإرهابيين. في هذا الموقف لا نجد أيًّا من النوعين مقبولًا، وهو السبب الذي لأجله ما زال الجدل محتدمًا في المجتمع حول التوازن المناسب بين محاربة الإرهاب وحماية الحريات المدنية. تنص الفرضية المنعدمة هنا على أن شخصًا ما ليس بإرهابي. وكما هو الحال في سياق الجرائم العادية، لا نريد أن نرتكب خطأً من النوع ١ ونرسل أبرياء إلى سجن خليج جوانتنامو. لكن في نفس الوقت، وفي عالم به أسلحة دمار شامل، يصير لجعل ولو إرهابي واحد حرًّا (خطأ من النوع ٢) نتائج كارثية بالمعنى الحرفي للكلمة. وذلك هو سبب احتجاز الولايات المتحدة الأمريكية – سواء أحببت هذا أم لم تفعل – لأشخاص يشتبه أنهم إرهابيون في خليج جوانتانامو، استنادًا إلى قرائن لا تكفي لإدانتهم في محكمة جنائية عادية.

ليس الاستدلال الإحصائي بضرب من ضروب السحر، ولا هو أمرًا لا تشوبه شائبة أو عيب، وإنما هو وسيلة ممتازة لإيجاد المنطق في العالم. فبه نستطيع أن نفهم الكثير من الظواهر الحياتية فهمًا عظيمًا؛ لأنه يجعلنا نقرر أيُّ تفسيرات تلك الظواهر أرجح. ومعظمنا لا يفتأ يفعل ذلك طوال الوقت (فمثلًا «عندما نرى طالبًا جامعيًّا مغشيًّا عليه في الأرض محاطًا بعبوات الجعة، نفكر أنه قد أسرف في الشراب» هي عبارةٌ أكثر منطقية من

عبارة «عندما نرى طالبًا جامعيًّا مغشيًّا عليه في الأرض محاطًا بعبوات الجعة، نفكر أنه قد سممه الإرهابيون»).

ولا يفعل الاستدلال الإحصائي سوى إضفاء صبغة رسمية على الأمر.

ملحق الفصل التاسع

حساب الخطأ المعياري للفارق بين المتوسطات

معادلة المقارنة بين متوسطين:

$$\frac{\bar{x}-\bar{y}}{\sqrt{\frac{s_x^2}{n_x}+\frac{s_y^2}{n_y}}}$$

البسط يشير إلى حجم الفارق بين المتوسطات

المقام يشير إلى الخطأ المعياري للفارق في المتوسطات بين العينتين

حيث

$\bar{x}$ = متوسط العينة x

$\bar{y}$ = متوسط العينة y

s_x = الانحراف المعياري للعينة x

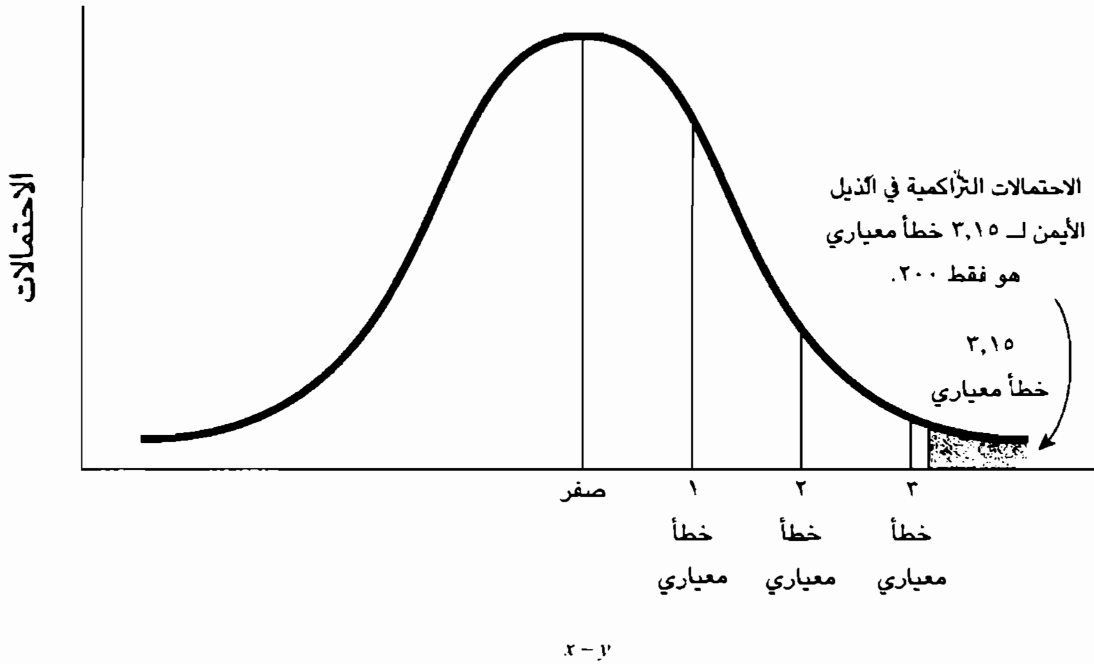
s_y = الانحراف المعياري للعينة y

n_x = عدد المشاهدات في العينة x

n_y = عدد المشاهدات في العينة y

تنص الفرضية المنعدمة على أن متوسطَي العينتين متطابقان. المعادلة السابقة قامت بحساب الفارق في المتوسطات نسبةً إلى حجم الخطأ المعياري للفارق في المتوسطات. مرة أخرى نعتمد بشدة على التوزيع الطبيعي. فإذا ما تطابقت بالفعل متوسطات المجتمعات التي أُخذت العينات منها، فلنا حينها أن نتوقع أن يقع الفارق بين متوسطات العينات في نطاق يقل عن خطأ معياري واحد في نسبة ٦٨ بالمائة من الحالات، وفي نطاق خطأين معياريين في ٩٥ بالمائة من الحالات، وهلم جرًّا.

في مثال دراسة التوحد التي أوردناها في هذا الفصل كان الفارق في المتوسطات بين العينتين ٧١,٦ سنتيمترًا مكعبًا بخطأ معياري قدره ٢٢,٧. وكانت نسبة الفارق الملاحظ ٣,١٥، بمعنى أن للعينتين متوسطين متباعدين بقدر ثلاثة أخطاء معيارية. وكما ذكرنا في هذا الفصل، تكون احتمالات وقوع متوسطين للعينتين بهذا التباعد، في ظل تطابق متوسطَي المجتمعين، هو أمر نادر الحدوث جدًّا جدًّا. وعلى وجه التحديد نقول إن احتمالات ملاحظة الفارق في المتوسطين البالغ ٣,١٥ أخطاء معيارية أو أكثر هي ٠٠٢.



٢-١.

الفارق في متوسطَي العينتين

اختبارات الفرضيات ذات الذيل وذات الذيلين

طرح علينا هذا الفصل فكرة استخدام العينات لاختبارِ إذا ما كان لاعبو كرة السلة المحترفون لهم نفس أطوال باقي مجتمع الذكور بشكل عام. لكنني أخفيت تفصيلة وحيدة. ففرضيتنا المنعدمة تنص على أن لاعبي كرة السلة الذكور لهم نفس متوسط بقية الرجال في المجتمع بشكل عام. لكن ما أخفيته هو أن لدينا فرضيتين بديلتين ممكنتين.

إحدى هاتين الفرضيتين البديلتين هي أن لاعبي كرة السلة المحترفين من الذكور لهم متوسط أطوال يختلف عن متوسط أطوال المجتمع كاملًا، فقد يكونوا أكثر منهم طولًا أو قصرًا. تلك كانت نفس المقاربة التي استخدمتها أنت وقت أن هبطتَ إلى الحافلة المخطوفة ووزنتَ ركابها كي تقرر إذا ما كان أولئك الركاب مشاركين في دراسة الحيوات المتغيرة أم لا. حينها كان باستطاعتك أن ترفض الفرضية المنعدمة، والتي تنص على أن أولئك الركاب هم بالفعل من المشاركين في الدراسة، إذا ما كان متوسط أوزان الركاب أعلى بكثير من متوسط مجتمع المشاركين كاملة أو أدنى منها بكثير (وهو ما كان بالفعل في تلك الحالة). أما الفرضية البديلة الثانية فتنص على أن لاعبي كرة السلة المحترفين من الذكور أكثر طولًا في المتوسط من بقية الرجال في ذلك المجتمع. في هذه الحالة تخبرنا خلفيتنا

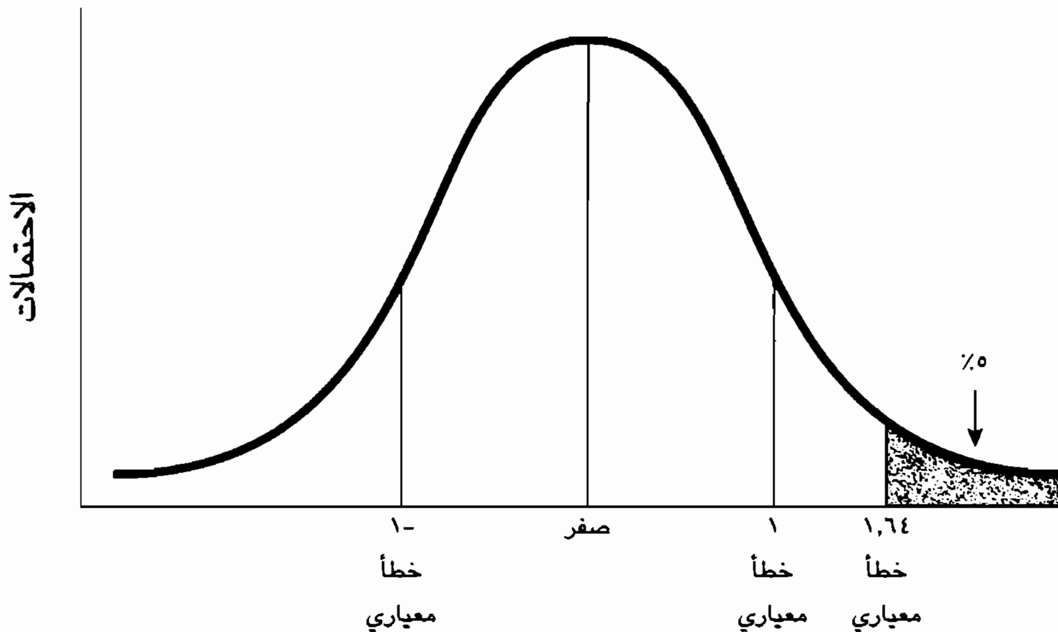
المعرفية أن لاعبي كرة السلة المحترفين لا يمكن أن يكونوا بحال من الأحوال أكثر قصرًا من بقية مجتمع الذكور. والفرق بين هاتين الفرضيتين البديلتين سيقرر إذا ما كنا سوف نجري اختبارًا للفرضية ذا ذيل واحد أو ذا ذيلين.

في كلتا الحالتين دعنا نفترض أننا سوف نجري اختبار دلالة عند مستوى ٠٫٠٥. لسوف نرفض الفرضية المنعدمة إذا ما لاحظنا فارقًا في الأطوال بين العينتين قد يحدث ٥ مرات أو أقل من أصل ١٠٠ فيكون لجميع الرجال أطوال متساوية. الأمر واضح حتى هذه النقطة.

لكن الأمور ستغدو مربكة فيما يلي. عندما تنص فرضيتنا البديلة على أن لاعبي كرة السلة المحترفين أكثر طولًا من غيرهم، حينها سنجري اختبار فرضية ذا ذيل واحد. سوف نقيس الفارق في المتوسطات بين في عينتنا التي تضم اللاعبين الذكور، وبين العينة الأخرى التي تضم رجالًا عاديين. نعلم أنه لو صحت فرضيتنا المنعدمة، فسوف نلاحظ حينها فارقًا يبلغ ١٫٦٤ خطأ معياري أو أكثر ٥ مرات من أصل كل ١٠٠ مرة. وسنرفض فرضيتنا المنعدمة إذا ما جاءت نتائجنا في النطاق الذي يبينه الشكل التالي.

الفارق في متوسطات العينات

(مقيسًا بالخطأ المعياري)



$x - y$

الفارق في متوسطَي العينتين

والآن دعنا نعُدْ إلى الفرضية البديلة الثانية التي ترى أن لاعبي كرة السلة المحترفين قد يكونون أكثر طولًا أو قصرًا من مجتمع الذكور بأسره. لن تختلف مقاربتنا العامة. مرة أخرى سوف نرفض فرضيتنا المنعدمة التي ترى أن لاعبي كرة السلة لهم نفس أطوال باقي المجتمع إذا ما حصلنا على تلك النتيجة ٥ مرات من أصل ١٠٠ مرة إذا لم يكن في الواقع اختلاف في الأطوال. الاختلاف يتمثل هنا في كوننا مضطرين للتعامل مع إمكانية كون لاعبي كرة السلة أقصر قامة من باقي الرجال في المجتمع. ومن ثم سوف نرفض الفرضية المنعدمة إذا ما كانت عينتنا المكونة من لاعبي كرة السلة الذكور لها متوسط أطوال أدنى من نظيره لدى عينة الرجال العاديين. وهذا يستلزم اختبار فرضية ذا ذيلين. ستتباين النقاط الفاصلة في رفض الفرضية المنعدمة؛ لأنه سيتعين علينا الآن الأخذ في الاعتبار إمكانية حدوث فارق كبير في متوسطات العينتين في كلا الاتجاهين الإيجابي والسلبي. وبشكل أكثر تحديدًا نقول إن نطاق رفض الفرضية المنعدمة قد انقسم بين ذيلين. وسوف نرفض أيضًا تلك الفرضية المنعدمة إذا ما خرجنا بناتج لا يحدث إلا في ٥ بالمائة من المرات أو أقل، إذا ما تساوت أطوال لاعبي كرة السلة مع أطوال بقية المجتمع، إلا أننا الآن نجد أمامنا طريقتين مختلفتين نرفض بهما تلك الفرضية المنعدمة.

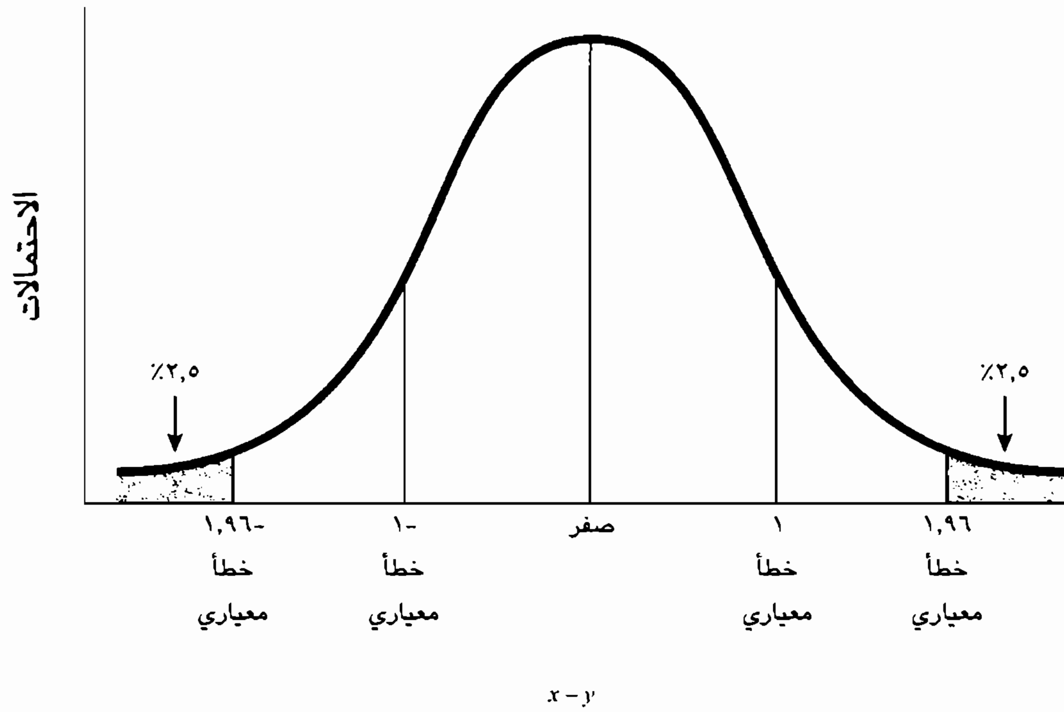
سوف نرفض الفرضية المنعدمة إذا كان متوسط أطوال عينة لاعبي كرة السلة المحترفين الذكور أكبر من نظيره لدى عينة الرجال العاديين، وهو الأمر الذي سنلاحظه ٢,٥ مرة فقط من أصل كل ١٠٠ مرة، إذا كان لاعبو كرة السلة لهم نفس أطوال الآخرين.

وسوف نرفض الفرضية المنعدمة إذا كان متوسط أطوال عينة لاعبي كرة السلة الذكور أدنى بكثير من نظيره لدى عينة الرجال العاديين، وهو الأمر الذي سنلاحظه فقط ٢,٥ مرة من أصل ١٠٠ مرة، إذا كان لاعبو كرة السلة لهم بالفعل نفس أطوال الآخرين.

وهذان الاحتمالان يجتمعان معًا فيتحولان إلى ٥ بالمائة كما يُظهر الشكل التالي.

الفارق في متوسطَي العينتين

(مقيسًا بالخطأ المعياري)



الفارق في متوسطَي العينتين

وحسن التقدير هو الذي نقرر من خلاله أيُّ الاختبارين، ذي الذيل الواحد أو الذيلين، هو الملائم لإجراء ذلك التحليل.

الفصل العاشر

استطلاعات الرأي

كيف لنا أن نعرف أن ٦٤ بالمائة من الأمريكيين يؤيدون عقوبة الإعدام (بخطأ استعيان قدره ± ٣ بالمائة)

في أواخر عام ٢٠١١ نشرت صحيفة النيويورك تايمز تحقيقًا إخباريًّا في صفحتها الأولى أعربت فيه عن «قلق بالغ وشكوك كبيرة حول المستقبل الذي ينتظر أمتنا»[١]. غاص ذلك التحقيق الإخباري في روح أمريكا كاشفًا عن موضوعات تتدرج من أداء إدارة أوباما إلى توزيع الثروة. إليك مقتطفات من آراء الأمريكيين في خريف عام ٢٠١١:

- نسبة صادمة قدرها ٨٩ بالمائة من الأمريكيين قالوا إنهم لا يثقون في أن الحكومة تفعل الصواب، وهو ما عُدَّ أعلى مستوًى مسجل من انعدام الثقة.
- قال ثلثا العامة إنه ينبغي أن توزع الثروة في البلاد بشكل أكثر تساوٍ.
- ثلاثة وأربعون بالمائة من الأمريكيين قالوا إنهم يؤيدون آراء حركة «احتلوا وول ستريت» في مجملها، وهي حركة غير منظمة بدأت قرب وول ستريت في مدينة نيويورك، وتفشت في العديد من المدن في جميع أرجاء الدولة*. وقال ستة وأربعون بالمائة من الأمريكيين، وهي نسبة أعلى قليلًا من سابقتها، إن آراء أولئك المشاركين في تلك الحركة «تعكس بشكل عام آراء معظم الأمريكيين».

---

* تعرف الحركة نفسها على موقعها الإلكتروني بالكلمات التالية: «حركة احتلوا وول ستريت هي حركة مدفوعة من الجماهير، بدأت في السابع عشر من سبتمبر من عام ٢٠١١ في ميدان الحرية بالمنطقة المالية بمانهاتن، ثم انتشرت فيما يزيد عن مائة مدينة بالولايات المتحدة وفيما يزيد عن ألف وخمسمائة مدينة على مستوى العالم. حركة احتلوا وول ستريت هي مقاومة لما للمصارف الضخمة والمؤسسات المتعددة الجنسيات من سلطة مستنزفة للعملية الديمقراطية، وكذلك للدور الذي لعبته وول ستريت في خلق الانهيار الاقتصادي الذي سبب أكبر كساد مالي حدث منذ عقود. لقد استمدت هذه الحركة إلهامها من الانتفاضات الشعبية التي اندلعت في مصر وتونس، والتي هدفت إلى فضح حقيقة أن نسبة الواحد بالمائة الأكثر ثراءً هم من يكتبون قواعد ذلك الاقتصاد العالمي الظالم الراهن لمستقبلنا».

- أثنى ستة وأربعون بالمائة من الأمريكيين على أداء باراك أوباما في منصبه – وأبدى ستة وأربعون بالمائة أيضًا عدم رضاهم عن ذلك الأداء.
- تسعة بالمائة فقط من العامة هم من أبدوا رضاءهم عن أداء الكونجرس لمهامه.
- رغم أنه لم يكن يفصلهم عن الانتخابات الأولية المؤهلة للرئاسة سوى شهرين، فإن نحو ثمانين بالمائة من مصوتي الحزب الجمهوري قالوا: «ما زال الوقت مبكرًا على تقرير من سيساندون من المرشحين».

تلك النسب المدهشة قد منحتنا أفكارًا ذات معنى عن آراء الأمريكيين قبل عام واحد من انطلاق السباق الرئاسي. لكن رغم هذا قد يتساءل أحدهم قائلًا كيف لنا أن نعرف كل هذا؟ كيف نستطيع أن نستخلص تلك النتائج الحاسمة عن توجهات مئات الملايين من المواطنين البالغين؟ وكيف لنا أن نعرف إذا ما كانت تلك النتائج الحاسمة دقيقة؟

الإجابة بالطبع هي أننا نجري استطلاعات للرأي، أو أن النيويورك تايمز وقناة الـ سي بي إس نيوز قد أَجْرَتَا استطلاع الرأي الذي تحصلنا منه على نتائج المثال السابق. (وكون مؤسستين إخباريتين تتعاونان معًا في مشروع كهذا، هو أول دليل على أن إجراء استطلاع رأي قومي سليم في منهجيته هو أمر ذو تكلفة ليست بالزهيدة). لا أشك في أنك لست غريبًا عن نتائج استطلاعات الرأي. لكن ما قد يغيب عن أذهاننا أن منهجية استطلاعات الرأي ما هي إلا شكل إضافي من أشكال الاستدلال الإحصائي. فاستطلاع الرأي أو (الاستقصاء) هو استدلال عن آراء مجتمعٍ ما استنادًا إلى آراء عبَّرَت عنها عينة منتقاة من ذلك المجتمع.

وقوة استطلاعات الرأي تنبع من نفس منبع مثالنا السابق، الذي هو مبرهنة النهاية المركزية. فإذا ما أخذنا عينة كبيرة ممثلة من المصوتين الأمريكيين (أو أي مجموعة أخرى)، يمكننا حينها أن نضع افتراضًا منطقيًّا مفاده أن العينة ستتشابه كثيرًا مع المجتمع الذي انتُقيت منه. وإذا ما كان نصف البالغين الأمريكيين يعارضون زواج المثليين، فلنا أن نتوقع أن نصف عينتنا الممثلة المؤلفة من ١٠٠٠ أمريكي سوف يرفضون هذا النوع من الزواج.

وبالعكس – والأكثر أهمية من وجهة نظر استطلاعات الرأي – إذا كانت لدينا عينة ممثلة مؤلفة من ١٠٠٠ أمريكي ممن لهم مشاعر معينة، كالستة والأربعين بالمائة غير الراضين عن أداء الرئيس أوباما، فمنها نستطيع أن نستدل على أن بقية المجتمع يشعر نفس الشعور على الأرجح. بل إننا نستطيع أن نحسب احتمالات انحراف نتائج العينة انحرافًا شديدًا عن التوجهات الحقيقية لذلك المجتمع. عندما تقرأ أن استطلاع رأيٍ ما له «هامش خطأ» يبلغ ± ٣ بالمائة، فهذا هو بالضبط نفس نوع فاصل الثقة البالغة نسبته ٩٥ بالمائة، والذي حسبناه في الفصل السابق. تعني «نسبة ٩٥ بالمائة من الثقة» أننا إذا أجرينا مائة استطلاع رأي مختلف على عينات من نفس المجتمع، فلنا أن نتوقع أن تكون

الإجابات التي نتحصل عليها من ٩٥ استطلاع من أصل المائة التي سنجريها على تلك العينة في نطاق ٣ بالمائة في أحد الاتجاهين اللذين تسير فيهما توجهات الناس الحقيقية. ففي سياق السؤال المتعلق بالرضا عن الأداء الرئاسي الذي طرح في استطلاع الرأي الذي أجرته النيويورك تايمز وقناة سي بي إس معًا، نستطيع أن نكون واثقين بنسبة ٩٥ بالمائة من أن النسبة الحقيقية من الأمريكيين غير الراضين عن أداء الرئيس أوباما تقع في نطاق ٤٦ بالمائة ± ٣ بالمائة، أو بين ٤٣ بالمائة و٤٩ بالمائة. لو كنت قد قرأت الخلاصة المختصرة لاستطلاع الرأي الذي أجرته النيويورك تايمز وسي بي إس (وأنصحك أن تفعل)، لوجدت فكرتها الأساسية فيما يلي: «من الناحية النظرية لن تختلف النتائج القائمة على كل عينة، في ١٩ حالة من أصل ٢٠، أكثر من ثلاث نسب مئوية في أي اتجاه من الاتجاهين اللذين رغبنا في معرفة ميل الأمريكيين لهما».

واحد من الفروق الأساسية بين استطلاعات الرأي وغيرها من وسائل أخذ العينات، هو أن العينة الإحصائية التي نهتم بها لن تكون متوسطًا (مثل ١٨٧ رطلًا)، وإنما نسبة مئوية أو حصة (على غرار ٤٧ بالمائة من المصوتين أو ٠٫٤٧). لكن ذلك الفرق وغيره من الفروق تختفي في الجوانب الأخرى من العملية. فعندما نحوز عينة كبيرة ممثلة (استطلاع الرأي)، سنتوقع أن نسبة المتجاوبين مع الاستطلاع في العينة – والذين يميلون إلى توجُّه معين (مثل التسعة بالمائة الذين يرون الكونجرس يبلي بلاءً حسنًا) – ستكون مساوية، تقريبًا، لنسبة جميع الأمريكيين الذين يميلون إلى نفس التوجه. لا يختلف هذا عن متوسط أوزان العينة المؤلفة من ١٠٠٠ من الذكور الأمريكيين الذي يساوي متوسط أوزان جميع الذكور الأمريكيين. لكننا رغم هذا نتوقع بعض التغاير في نسب مَن يقرون أداء الكونجرس من عينة لأخرى، تمامًا كما كنا نتوقع أن تتغاير متوسطات أوزان العينات العشوائية المختلفة، والتي يبلغ قوام الواحدة منها ١٠٠٠ رجل. لو أجرت النيويورك تايمز والـ سي بي إس استطلاع رأي ثانيًا – طارحتين نفس الأسئلة على عينة جديدة قوامها ١٠٠٠ أمريكي بالغ – فمن المستبعد كثيرًا أن تتطابق الأجوبة مع أجوبة الاستطلاع الأول. لكن مع ذلك لا ينبغي أن تنحرف إجابات الاستطلاع الثاني انحرافًا شديدًا عن إجابات سابقه. (وبالعودة لمثال ذكرناه سابقًا، تَذكَّرْ أنك إن تذوَّقتَ ملعقة من الحساء ثم قلَّبته في قِدره وتذوَّقتَ ملعقة أخرى، فلن تختلف تلك الملعقة في مذاقها عن سابقتها). والخطأ المعياري هو الذي ينبئنا بمدى الانحراف الذي نتوقعه في نتائج عينة عن أخرى، والذي ينطبق في هذه الحالة على الفارق بين استطلاع رأي وآخر.

تختلف معادلة حساب الخطأ المعياري لحصة أو نسبة مئوية ما اختلافًا طفيفًا عن المعادلة التي ذكرناها سابقًا، لكن منطقهما متطابق جدًّا. فالخطأ المعياري لأي عينةٍ مأخوذةٍ عشوائيًّا يساوي ؛ $\sqrt{p(1-p)/n}$ حيث p هو حصة المتجاوبين مع الاستطلاع

الذين يعبِّرون عن وجهة نظر معينة، و (1-p) هو حصة المتجاوبين ذوي وجهة نظر مغايرة، وn هو إجمالي عدد المتجاوبين في العينة. سوف ترى أنه كلما كبر حجم العينة انخفض قدر الخطأ المعياري بما أن n هو المقام. وكذلك فإن الخطأ المعياري يميل لأن يصغر عندما يتباعد الـ p والـ (1-p) كثيرًا. فمثلًا يصير الخطأ المعياري أصغر في استطلاع رأي يشترك فيه ٩٥ بالمائة من المتجاوبين في وجهة نظر واحدة، عنه في استطلاع رأي آخر تنقسم به الآراء النصف للنصف. ما الأمر إلا محض حسابات رياضية، فإذا كانت (٠٥) (٩٥.) = ٠٤٧. بينما (٥.) (٥.) = ٢٥. فإن الرقم الأصغر في المعادلة يؤدي إلى رقم أصغر في الانحراف المعياري.

لنفترض، على سبيل المثال، أن «استطلاع مسح اقتراع ناخبين» بسيطًا قد أجري على خمسمائة من المصوتين المثلين يوم الانتخابات، فوُجد أن ٥٣ بالمائة قد صوتوا لصالح المرشح الجمهوري، وأن ٤٥ بالمائة قد صوتوا لصالح المرشح الديمقراطي، وصوَّت ٢ بالمائة لصالح مرشح حزب ثالث. لو جعلنا المرشح الجمهوري هو الحصة التي نهتم بها، فإن الخطأ المعياري لمسح اقتراع الناخبين ذلك سيكون:

$$\sqrt{(.53)(1-.53)/500} = \sqrt{(.53)(.47)/500} = \sqrt{.25/500} = \sqrt{.0005} = .2236$$

طلبًا للتبسيط سوف نقوم بتقريب الخطأ المعياري إلى ٠٢. وما هذا حتى الآن سوى رقم. لكن دعنا نتفكر في سبب أهمية ذلك الرقم. ولنفترض أن الاستطلاع انتهى وكنت تعمل لدى محطة تلفزيونية ترغب في إعلان الفائز بسباق الرئاسة قبل الإعلان عن النتائج النهائية. ولأنك الآن الحَكَم الرسمي على البيانات في تلك المحطة التلفزيونية (حيث إنك قد قرأت ثلثي هذا الكتاب)، وسألتْك منتجة البرنامج عما إذا كان من الممكن «إعلان نتيجة السباق» استنادًا إلى مسح اقتراع الناخبين.

تشرح لها أن الإجابة عن هذا السؤال تتوقف على مدى الثقة التي يرغب مسئولو الشبكة أن يكونوا عليها عندما يقوموا بهذا الإعلان – أو بشكل أكثر تحديدًا مدى المخاطرة التي يقبلون تحملها لأن يكونوا مخطئين. تذكر أن الخطأ المعياري يجعلنا نتنبأ بوتيرة وقوع حصة عينتنا (مسح اقتراع الناخبين) مقاربة بشكل معقول لحصة المجتمع الحقيقي (نتائج الانتخابات). إننا نعلم أن لنا في ٦٨ بالمائة تقريبًا من الحالات أن نتوقع حصة العينة – الثلاثة والخمسين بالمائة من المصوتين الذين قالوا إنهم قد صوتوا للجمهوريين في هذه الحالة – سوف تقع في نطاق خطأ معياري واحد من الإحصاء النهائي الحقيقي. ونتيجة لهذا تخبر أنت منتجة البرنامج «بثقة مقدارها ٦٨ بالمائة» أن عينتك، التي تُظهر حصول الجمهوريين على ٥٣ بالمائة من الأصوات ± ٢ بالمائة، أو ما بين ٥١ و٥٥ بالمائة، قد أثبتت فوز الجمهوريين بالسباق. لكن في الوقت نفسه يظهر نفس مسح اقتراع العينات أن المرشح الديمقراطي قد حصل على ٤٥ بالمائة من الأصوات. إذا ما افترضنا أن إحصاء

أصوات المرشح الديمقراطي له نفس الخطأ المعياري (وهو تبسيط سوف أشرحه بعد قليل)، فنستطيع أن نقول بثقة قدرها ٦٨ بالمائة إن عينة مسح الاقتراع، التي تظهر حصول الديمقراطيين على ٤٥ بالمائة ± ٢ بالمائة، أو ما بين ٤٣ و٤٧ بالمائة، قد عبَّرت عن الإحصاء الحقيقي للأصوات التي حازها الديمقراطيون. وطبقًا لتلك الحسابات فقد فاز الجمهوريون.

هنا يهرع أفراد قسم الجرافيك كي يرسموا صورة ثلاثية الأبعاد كي تسطع على الشاشة ليراها المشاهدون تنص على ما يلي:

الجمهوريون ٥٣٪

الديمقراطيون ٤٥٪

المستقلون ٢٪

(هامش الخطأ ٢٪)

في البداية تُسَرُّ المنتجة وتتحمس، والسبب الأكبر لهذا هو تلك الصورة الملونة الثلاثية الأبعاد التي تدور حول نفسها على الشاشة. لكنك عندما تشرح لها أنه في ٦٨ حالة تقريبًا من أصل ١٠٠ سوف تقع نتائج مسح الاقتراع الذي أجريته في نطاق خطأ معياري واحد من النتائج الحقيقية للانتخابات، هنا سوف تقوم منتجة البرنامج، وهي التي سبق وأجبرتها المحكمة مرتين على أن تخضع لبرنامج علاجي للتحكم في الغضب، بالإشارة إلى عملية حسابية واضحة، ألا وهي أنه في ٣٢ مرة من أصل ١٠٠ لن يقع مسح الاقتراع الذي أجريته في نطاق خطأ معياري واحد من النتائج الحقيقية للانتخابات. إذن ماذا سيحدث؟

حينها تشرح لها أن هناك احتمالين: (١) أن المرشح الجمهوري يمكن أن يكون قد حصد أصواتًا أكثر من التي توقعها استطلاع الرأي، وفي تلك الحالة كانت ستظل تلك النتيجة التي أعلنتها صحيحة من حيث اسم الفائز بالسباق. أو (٢) هناك احتمال كبير نوعًا ما أن يكون المرشح الديمقراطي قد حاز أصواتًا تتعدى بكثير ما أشار إليه استطلاع الرأي الذي أجريته، وفي هذه الحالة ستكون صورتك الملونة الدوارة تلك قد أعلنت المرشح الخطأ فائزًا.

حينها تطوح منتجة البرنامج بقدح قهوتها إلى نهاية الغرفة، وتستخدم عبارات عديدة تخرق اتفاقية وقف عقوبتها فتصرخ: «كيف لنا أن نكون [كلمة تحذفها الرقابة] متيقنين من أننا نحوز النتائج [كلمة تحذفها الرقابة] الصحيحة؟»

ولأنك زعيم الإحصاء، فقد وضحت لها أنك لن تستطيع أن تتيقن من أي نتيجة إلى أن ينتهي عد الأصوات. لكنك مع ذلك تستطيع أن تزودها بفاصل ثقة قدره ٩٥ بالمائة. في هذه الحالة ستخطئ صورتك الملونة الثلاثية الأبعاد فيما متوسطه ٥ مرات من أصل ١٠٠.

هنا تشعل المنتجة سيجارتها وتبدو عليها ملامح الارتياح. وتقرر أنت ألا تذكِّرها بحظر التدخين في أماكن العمل لأن ذلك أدى في المرة السابقة إلى نتائج كارثية. لكنك مع ذلك قررت أن تطلعها على بعض الأنباء السيئة، وهي أن الوسيلة الوحيدة التي تستطيع بها المحطة أن تحوز المزيد من الثقة في نتائج استطلاعها هي من خلال توسيع «هامش الخطأ». وعندما تفعل ذلك لن يعود هناك حسم لنتيجة الانتخابات. هنا ستعرض على رئيستك تلك هذه الصورة الجديدة الخلابة:

الجمهوريون ٥٣٪

الديمقراطيون ٤٥٪

المستقلون ٢٪

(هامش الخطأ ٤٪)

لقد أخبرتنا مبرهنة النهاية المركزية سابقًا أن ٩٥ بالمائة من حصة العينة سوف تقع في نطاق خطأين معياريين من حصة المجتمع الحقيقي (والتي هي ٤٪ في هذه الحالة). ومن ثم فإن أردنا أن نكون أكثر ثقة في نتائج استطلاع الرأي، فعلينا أن نخفض من سقف توقعاتنا. كما تظهر الصورة السابقة (دونما ألوان أو أبعاد ثلاثية)، يمكن أن تعلن المحطة بثقة مقدارها ٩٥ بالمائة أن المرشح الجمهوري قد حصد ٥٣ بالمائة من الأصوات ± ٤، أو ما بين ٤٩ و٥٧ بالمائة من الأصوات. لكن في نفس الوقت قد حصد المرشح الديمقراطي ٤٥ بالمائة ± ٤ بالمائة، أو ما بين ٤١ و٤٩ بالمائة من عدد الأصوات.

نعم، أمامك الآن مشكلة جديدة. فعند مستوى ٩٥ من الثقة لا تستطيع أن تتنبأ إمكانية تعادُل المرشحين بحصول كل منهما على نسبة ٤٩ بالمائة. تلك هي المساومة المحتومة، فالسبيل الوحيد للإنعام في التيقن من توافق توقعات استطلاع الرأي الذي تجريه مع نتائج الانتخابات من دون البيانات الجديدة، هو أن تكون أكثر تحفظًا في توقعاتك. دعنا نتأمل سياقًا آخر غير إحصائي. لنفترض أنك قلت لصديق لك إنك «واثق جدًّا» أن توماس جيفرسون كان الرئيس الثالث أو الرابع للولايات المتحدة. كيف لك أن تغدو أكثر ثقة في معلوماتك التاريخية؟ بأن تكون أقل تحديدًا. فأنت «متأكد جدًّا» أن توماس جيفرسون كان من بين الرؤساء الخمسة الأوائل للولايات المتحدة الأمريكية.

تطلب منك منتجة البرنامج أن تطلب لنفسك فطيرة بيتزا وتجهز نفسك لأن تبيت ليلتك في العمل. هنا يحالفك حظ إحصائي حسن. فتأتيك على مكتبك نتائج مسح الاقتراع لعينة قوامها ألفا مصوت. تظهر النتائج ما يلي: الجمهوريون (٥٢ بالمائة) والديمقراطيون (٤٥ بالمائة) والمستقلون (٣ بالمائة). هنا يتملك منتجة البرنامج شديد الغضب؛ وذلك لأن استطلاع الرأي قد رجح ضيق الفجوة بين المرشحين مصعبًا عليكم مهمة إعلان نتائج السباق الرئاسي في وقت قريب. لكن تمهل! هنا تقوم بعمل (بطولي)

وتشير إلى أن حجم العينة (٢٠٠٠) يزيد عن حجم عينة الاستطلاع الأول بأربعة أمثال. ونتيجة لهذا سوف يتقلص الخطأ المعياري بشكل كبير. وهكذا يكون الخطأ المعياري الجديد للمرشح الجمهوري $\sqrt{.52(.48)/2000}$ أي ٠١.

إذا ما اطمأنت المنتجة إلى نسبة الثقة البالغة ٩٥ بالمائة، فحينها تستطيع إعلان المرشح الجمهوري فائزًا. مع الخطأ المعياري الجديد الذي ظهر لك ٠١. يصير فاصل الـ ٩٥ بالمائة من الثقة للمرشحين كما يلي: الجمهوريون: ٥٢± ٢ أو ما بين ٥٠ و٥٤ بالمائة من إجمالي الأصوات، الديمقراطيون: ٤٥ ± ٢، أو ما بين ٤٣ و٤٧ بالمائة من إجمالي الأصوات. لم يعد هناك تداخل بين فاصلي الثقة. حينها تستطيع أن تظهر على الهواء مباشرة وتتنبأ بفوز المرشح الجمهوري، وفي ٩٥ حالة من أصل ١٠٠ سيكون توقعك صحيحًا*.

لكن هذه الحالة لا تقف عند ذلك الحد، وإنما تتعداه إلى ما هو أفضل. تخبرنا مبرهنة النهاية المركزية أنه في ٩٩,٧ بالمائة من الحالات ستقع حصة العينة في نطاق ثلاثة أخطاء معيارية من حصة كامل المجتمع الذي انتُقيت منه. في مثال الانتخابات هذا تتمثل فواصل الثقة ذات النسبة المئوية البالغة ٩٩,٧ لكلا المرشحين كما يلي: الجمهوري ٥٢ ± ٣ بالمائة أو ما بين ٤٩ و٥٥ بالمائة، الديمقراطي ٤٥ ± ٣ بالمائة أو ما بين ٤٢ و٤٨ بالمائة. لو أعلنتم فوز المرشح الجمهوري فستتضاءل للغاية احتمالات فصلك ومنتجة برنامجك من عملكما بالمحطة، وذلك بفضل عينتك الجديدة المؤلفة من ألفَي مصوت.

عليك أن تلاحظ أن كبر حجم العينة يؤدي إلى تقليص حجم الخطأ المعياري، وهو الأمر الذي يجعل استطلاعات الرأي القومية المتسعة النطاق دقيقةً بشكل مذهل. لكن على الجهة المقابلة، كلما صغرت العينة اتسع نطاق الأخطاء المعيارية، ومن ثم اتسع نطاق فاصل الثقة (أو «هامش خطأ الاستعيان» طبقًا لمصطلحات استطلاعات الرأي). في استطلاع النيويورك تايمز وسي بي إس، كُتب بأحرف صغيرة لا تلاحَظ أن هامش الخطأ للأسئلة المتعلقة بالانتخابات المؤهلة للترشيح عن الحزب الجمهوري ذو نقاط تبلغ نسبتها ٥ بالمائة مقارنة ببقية الأسئلة التي وردت في الاستطلاع. تلك الأسئلة لم تطرح إلا على مَن قالوا عن أنفسهم إنهم قد صوتوا في الانتخابات التمهيدية والمجمع الانتخابي، ولذلك فقد هوى حجم العينة لتلك المجموعة الفرعية من الأسئلة إلى ٤٥٥ (مقارنة بـ ١٦٥٠ بالغًا هم قوام عينة استطلاع الرأي).

---

* في ما نسبته ٥ بالمائة من الحالات، لنا أن نتوقع أن يكون تعداد الأصوات الحقيقي للمرشح الجمهوري خارج فاصل الثقة. في مثل تلك الحالات يصير التعداد الحقيقي للأصوات التي حازها أقل من ٥٠ بالمائة أو أكبر من ٥٤ بالمائة. لكن رغم ذلك لو حصل على أكثر من ٥٤ بالمائة لن تكون محطتك قد أخطأت. (فإنك لم تفعل سوى أن قلصت هامش فوزه). ونتيجة لذلك ستقدر احتمالات أن يتسبب استطلاع الرأي ذلك في أن يجعلك تخطئ في إعلان المرشح الجمهوري فائزًا بـ ٢,٥ بالمائة.

إنني كما هي العادة قد قمت بتبسيط الكثير من الأمور في هذا الفصل. لعلك قد لاحظت ذلك في مثال الانتخابات الذي أوردتُه للتوِّ، فلسوف يكون للمرشحين الجمهوري والديمقراطي كلٍّ منهما الخطأ المعياري الخاص به. فلتتأمل المعادلة مرة أخرى: $SE = \sqrt{p(1-p)/n}$ حجم العينة n هو نفسه لدى المرشحين، أما الـ p و (1-p) فيتباينان قليلًا. في مسح الاقتراع الثاني (ذو العينة المكونة من ٢٠٠٠ مصوت) يكون الخطأ المعياري للمرشح الجمهوري $\sqrt{.52(.48)/2000} = .01117$ وللمرشح الديمقراطي $SE = \sqrt{.45(.55)/2000} = .01112$ وبالطبع هذان الرقمان متطابقان من الناحية العملية. ولذلك السبب فقد انتهجت المنهج الشائع، والذي يتمثل في اتخاذ أعلى الخطأين المعياريين واستخدامه لكلا المرشحين. وهو الأمر الذي من شأنه أن يجعلنا أكثر حيطة فيما يتعلق بفواصل الثقة.

لكن الكثير من استطلاعات الرأي القومية التي تطرح أسئلة متعددة سوف تمضي متخذة خطوة إضافية. في حالة استطلاع النيويورك تايمز والـ سي بي إس، سيكون الخطأ المعياري مختلفًا من الناحية التقنية لكل واحد من أسئلته استنادًا إلى إجابة هذا السؤال. فمثلًا لا بد أن الخطأ المعياري للنتيجة التي مفادها أن ٩ بالمائة من العامة راضون عن أداء الكونجرس لمهامه، سيكون أدنى من الخطأ المعياري للنتيجة التي تفيد بأن ٤٦ بالمائة من العامة راضون عن أداء الرئيس أوباما لمهامه؛ لأن ٠٩. × (٩١.) أقل من ٤٦. × (٥٤.) ٠٨١٩. في مقابل ٢٤٨٤. (سوف يَرِدُ شرح المنطق من وراء هذه المعادلة في ملحق الفصل).

ولأنه من المربك والمرهق أن يكون لدينا خطأ معياري مختلف لكل واحد من الأسئلة، فإن الاستطلاعات من هذا النوع ستفترض بشكل تلقائي أن حصة العينة لكل سؤال تبلغ ٥. (أو ٥٠ بالمائة) – مولدة أكبر خطأ معياري ممكن لأي عينة كانت – ومن ثم تبني ذلك الخطأ المعياري لحساب هامش خطأ الاستعيان للاستطلاع كاملًا*.

إن استطلاعات الرأي التي تجرى بشكل صحيح هي أدوات غريبة. يذكر فرانك نيوبورت رئيس تحرير مؤسسة جالوب أن استطلاعًا لرأي ألف شخص يمكن أن يمنحنا رؤية دقيقة لتوجهات الدولة بأكلمها. كلامه صحيح من الناحية الإحصائية. لكن كي نخرج بتلك النتائج الدقيقة ذات الدلالات، لا بد أن نجري استطلاع الرأي بشكل صحيح، وكذلك نترجم نتائجه على الوجه الصحيح، وكلاهما أمران سهل أن نقولهما، صعب أن ننجزهما. فعندما نحسب الأخطاء المعيارية لا نخرج بالنتائج الخطأ من الحسابات الرياضية الخطأ. وإنما تنتج النتائج الخطأ عن العينات المتحيزة أو الأسئلة الخطأ أو كلتيهما معًا. إن مقولة

* معادلة حساب الخطأ المعياري تلك التي أوردتها هنا تفترض أن الاستطلاع قد أجري على عينة عشوائية منتقاة من المجتمع. قد تنحرف المؤسسات العاملة في إجراء مثل تلك الاستطلاعات عن ذلك المنهج في أخذ العينات، في هذه الحالة تتغير معادلة حساب الخطأ المعياري تغيرًا طفيفًا. لكن أساس المنهجية يظل كما هو.

«تدخل قمامة، فتخرج قمامة» تنطبق تمام الانطباق على عملية أخذ عينات من الرأي العام. فيما يلي الأسئلة المنهجية الرئيسية التي يجدر بالمرء أن يطرحها عندما يجري استطلاعًا للرأي، أو عندما يراجع أعمال الآخرين.

هل هذه عينة دقيقة متخذة من مجتمع الأشخاص الذين نحاول قياس آرائهم؟ في الفصل السابع ناقشنا الكثير من التحديات الشائعة ذات الصلة بالبيانات. لكن رغم ذلك سوف أقوم بالإشارة مجددًا إلى خطورة الانحياز الانتقائي، والانتقاء الذاتي على وجه التحديد. فأي استطلاع للرأي يقوم على أفراد يختارون الانضمام إلى العينة، كبرامج الإذاعة التي تطلب اتصالات المواطنين أو الاستقصاءات التي تجرى عبر شبكة الإنترنت، سوف يجمع فقط آراء أولئك الذين بذلوا الجهد كي يصرحوا بتلك الآراء. وعادة ما يكون هؤلاء الأشخاص ذوي مشاعر قوية تجاه قضية ما، أو أشخاصًا تَصادَف أنهم يحوزون الكثير من أوقات الفراغ. وأيٌّ من هاتين المجموعتين لا تعتبر ممثلة لآراء غالبية العامة. ذات مرة شاركت في أحد برامج الإذاعة كضيف عبر الهاتف. ثم اتصل أحد المستمعين بالبرنامج معربًا على الهواء مباشرة بكل قوة عن أن آرائي «خطأ جدًّا»، لدرجة أنه توقف بسيارته على جانب الطريق السريع وبحث عن هاتف عام كي يهاتف البرنامج ويصرح برفضه لما قيل. وإنني أحب أن أعتقد أن أولئك المستمعين الذين لم يوقفوا سياراتهم جانب الطريق السريع لم يشعروا كما شعر.

وأي منهجية في جمع الآراء تستبعد قسمًا من المجتمع بشكل نظامي هي كذلك عرضة للانحياز. فمثلًا قد هيأت الهواتف المحمولة وسطًا حاضنًا لتعقيدات منهجية جديدة. ومؤسسات استطلاع الآراء المحترفة تبذل الغالي والنفيس في سبيل أخذ آراء عينة ممثلة للمجتمع المنشود. ولقد أجري استطلاع رأي النيويورك تايمز والـ سي بي إس من خلال اتصالات هاتفية أجريت على مدى ستة أيام لـ ١٦٥٠ من البالغين، قال ١٤٧٥ منهم إنهم ممن يحق لهم التصويت.

ليس بإمكاني سوى أن أخمن بقية المنهجية التي اتبعوها، لكن معظم استطلاعات الرأي الاحترافية تستخدم تنويعات من الأساليب التالية. للتأكد من أن البالغين الذين أجابوا تلك الاتصالات الهاتفية هم ممثلون لذلك المجتمع، تبدأ العملية باحتمالات ما – هي تنويع على عملية انتقاء كرة زجاجية من جَرَّة. فيقوم الحاسوب بانتقاء عشوائي لمجموعة من توليفات أرقام الهواتف ذات الخطوط الأرضية. (توليفة الرقم هي رمز المنطقة مضافًا إليه الأرقام الثلاثة الأولى من رقم الهاتف). وبالانتقاء العشوائي من بين ٦٩٠٠٠ توليفة أرقام هواتف أرضية في الدولة، كلٌّ نسبةً إلى نصيبه من إجمالي عدد أرقام الهواتف، سيحظى هذا الاستقصاء على الأرجح وبشكل عام بتوزيع جغرافي ممثل للمجتمع. تقول المقدمة التعريفية للاستقصاء والمكتوبة بالخط الصغير: «توليفات الأرقام تلك اختيرت

بشكل يضمن تمثيل كل منطقة من مناطق الدولة نسبةً لنصيبها من إجمالي أرقام الهاتف بالدولة». فقد قام الكمبيوتر بإضافة أربعة أرقام عشوائية لكل توليفة أرقام اختيرت. ونتيجة هذا سوف تندرج كل الأرقام المدرجة وغير المدرجة في القائمة النهائية للأرقام المنزلية التي سوف يتم الاتصال بها. وتضمَّن الاستقصاء أيضًا «اتصالًا عشوائيًّا بأرقام هواتف محمولة».

وفي كل اتصال يجرى كان القائمون على الاستقصاء يتأكدون من أن المتجاوب معهم شخص بالغ، فحتى لو أجاب الهاتف طفل كانوا يطلبون منه أن ينادي لهم أي شخص بالغ موجود في البيت الآن. لقد نُقِّحت تلك العملية لأجل إنتاج عينة من المتجاوبين تُشابِه مجتمع البالغين من حيث العمر والجنس. والأكثر أهمية من ذلك أن مُجري الاتصالات سوف يحاول إجراء العديد منها في ساعات مختلفة من الليل والنهار، بهدف أن يتمكن من الوصول إلى كل الأرقام المختارة. وهذه المحاولات المكررة – التي قد تبلغ عشرًا أو اثنتي عشرة مكالمة لنفس الرقم – هي جزء مهم من العملية وأساسي في الحصول على عينة غير منحازة. بالطبع سيكون من الأسهل والأقل تكلفة أن تجرى اتصالات عشوائية بأرقام مختلفة، حتى يستكمل العدد المطلوب لبناء تلك العينة الكبيرة المؤلفة من أشخاص بالغين قد أجابوا هواتفهم، وأجابوا كذلك على تلك الأسئلة التي طرحت عليهم. لكن مع ذلك ستكون تلك العينة أيضًا منحازة، منحازة لأولئك الذين يكثر مكوثهم في البيوت ويجيبون هواتفهم، كالعاطلين عن العمل وكبار السن وقبيلهم. لن يمثل ذلك مشكلة إذا ما كنت ترغب في أن تخرج نتائج استطلاعك على هذا النحو: بلغ معدل الرضا عن أداء الرئيس أوباما ٤٦ بالمائة بين العاطلين عن العمل والمسنين وغيرهم ممن يحبون الرد على الاتصالات الهاتفية العشوائية.

يعتبر معدل التجاوب مع استطلاع الرأي واحدًا من المؤشرات على صلاحيته؛ بمعنى كم قَدْرُ من أكملوا الإجابة على استطلاع الرأي أو الاستقصاء من بين من اختيروا وتم الاتصال بهم؟ وانخفاض معدل المتجاوبين قد يكون إشارة منذرة بعينة منحازة. فكلما زاد عدد من فضلوا عدم الاشتراك في استطلاع الرأي هذا أو من تعذر الاتصال بهم، زادت احتمالات كون هذه المجموعة الكبيرة تختلف اختلافًا ملموسًا عن أولئك الذين أجابوا عن الأسئلة. يمكن أن يصاب القائمون على استطلاعات الرأي بما يطلق عليه «انحياز عدم الاستجابة» إذا ما قاموا بتحليل البيانات المتاحة عن الأشخاص الذين تعذر عليهم التواصل معهم. هل يقطنون منطقة بعينها؟ هل يرفضون الرد على الاتصالات الهاتفية لسبب بعينه؟ هل من المحتمل أن يكونوا منتمين إلى مجموعة عرقية أو قومية أو تنتمي إلى مستوًى معين من الدخل المالي؟ هذا النوع من التحليل قادر على تحديد إذا ما كان معدل التجاوب الضعيف سيؤثر على نتائج استطلاع الرأي.

هل وضعتَ الأسئلة بطريقة تستخرج معلومات دقيقة حول الموضوع الذي تهتم به؟ إن الْتِماس معرفة الرأي العام يتطلب ما هو أكثر من قياس نتائج الاختبارات أو وضع المتجاوبين على موازين لتحديد أوزانهم. فنتائج أي استبيان يمكن أن تكون حساسة للغاية للطريقة التي يُسأل بها السؤال. لنتأمل مثالًا يبدو بسيطًا: كم من الأمريكيين يناصرون عقوبة الإعدام؟ كما يقول عنوان الفصل، فإن غالبية الأمريكيين يؤيدون تلك العقوبة. ذكرت مؤسسة جالوب أنه ومنذ عام ٢٠٠٢ أعلن ما يزيد عن ٦٠ بالمائة من الأمريكيين أنهم يفضلون تنفيذ عقوبة الإعدام في حق من أدينوا بالقتل. لقد تأرجحت نسبة الأمريكيين المناصرين لتلك العقوبة القصوى في إطار نطاق ضيق، يتراوح ما بين ٧٠ بالمائة في أعلى مستوياته عام ٢٠٠٣، إلى ٦٤ بالمائة في أدنى مستوياته التي تحققت في أزمنة مختلفة. إن بيانات استطلاع الرأي واضحة تنص على أن الأمريكيين يناصرون عقوبة الإعدام وبمعدل كبير.

أو لا يفعلون. فمناصرة الأمريكيين لعقوبة الإعدام تتراجع كثيرًا عندما يكون بديلها السجن مدى الحياة دون فرصة للإفراج المشروط. في عام ٢٠٠٦ خلص استطلاع رأي أجرته مؤسسة جالوب إلى أن ٤٧ بالمائة فقط من الأمريكيين يرون عقوبة الإعدام الجزاء المناسب لجريمة القتل، في مقابل ٤٨ بالمائة فضلوا عليها السجن مدى الحياة[٢]. ليست تلك بحقيقة إحصائية تافهة ترفِّه بها عن ضيوفك في إحدى حفلات العشاء التي تقيمها، فهي تعني أنه لم تعد الغالبية تناصر العقوبة القصوى عندما يتاح السجن مدى الحياة دونما إفراج مشروط كبديل معقول. وعندما نلتمس التعرف على الرأي العام، تكون صياغة الأسئلة واختيار المفردات أمرًا ذا أهمية كبيرة.

عادة ما يستغل السياسيون هذه الظاهرة عندما يستخدمون استطلاعات الرأي وجماعات التركيز في اختيار «الكلمات التي تخدم الغرض». فمثلًا يميل المصوتون لمناصرة «الإعفاء الضريبي» أكثر من «التخفيض الضريبي»، رغم أن العبارتين تصفان نفس الشيء. وكذلك لا يبالي المصوتون بالـ «التغير المناخي» كما يبالون بالـ «الاحتباس الحراري»، حتى لو كان الاحتباس الحراري هو أحد صور التغير المناخي. من الواضح أن السياسيين يحاولون التلاعب بأصوات الناخبين عن طريق اختيار كلمات غير محايدة. فلو اعتبرنا القائمين على استطلاعات الرأي وسطاء أمينين يخرجون علينا بنتائج شرعية، فعليهم أن يحذروا المفردات التي هي قابلة لأن تؤثر على دقة المعلومات التي جمعوها. وبالمثل فلو قورنت الإجابات على مر الزمن – كآراء المستهلكين عن الاقتصاد اليوم مقارنة بآرائهم قبل عام مضى – فلا بد أن تتطابق الأسئلة في المرحلتين أو تتشابه كثيرًا.

مؤسسات استطلاع الرأي مثل جالوب عادة ما تجري «اختبار العينة المشطورة»، والذي فيه تختبر تنويعات على سؤال واحد عن عينات مختلفة لقياس كم تؤثر التغييرات

الطفيفة في الصياغة على إجابات الخاضعين للاستقصاء. تمثل هذه الإجابات بالنسبة إلى الخبراء، مثل فرانك نيوبورت العامل بمؤسسة جالوب، بيانات ذات معنًى حتى عندما تبدو متضاربة[٣]. وحقيقة أن توجُّه الأمريكيين نحو العقوبة القصوى يتغير بشكل دراماتيكي عندما يتاح خيار السجن مدى الحياة دون إفراج مشروط، ينبئنا بشيء مهم. يقول نيوبورت إن الفكرة الأهم هنا هي أن تَعرِض نتائج أي استطلاع رأي كانت في سياقها. فليس هناك سؤال واحد أو استطلاع رأي وحيد قادرًا على أن يستوعب الرأي العام تجاه قضية ما بكامل عمقه.

هل يقول المتجاوبون مع استطلاع الرأي الحقيقة؟ إن استطلاعات الرأي هي أشبه بالمواعدة عبر الإنترنت: لا سبيل للتوثق من المعلومات المتاحة. فإننا نعلم أن الناس يعتِّمون على الحقيقة، خاصة عند الإجابة عن أسئلة محرجة أو حساسة. فقد يبالغ المتجاوبون فيما يجنون من دخل أو في عدد مرات ممارستهم للجنس في الشهر الواحد. وقد لا يعترفون أنهم صوتوا في الانتخابات. وقد يترددون في التصريح بآرائهم التي لا تلقى قبولًا شعبيًّا أو مجتمعيًّا. لكل تلك الأسباب نجد أنه حتى أكثرُ استطلاعات الرأي المصممة بعناية تعتمد على نزاهة إجابات الخاضعين لها.

تعتمد استطلاعات الآراء المتعلقة بالانتخابات بشكل جوهري على الفصل بين من سيصوتون يوم الانتخابات ومن لن يصوتوا. (فإذا ما كنا نحاول التوصل للفائز المحتمل بالانتخابات، فلن نبالي بآراء مَن لا يَنتوون التصويت). وعادة ما يقول الناس إنهم سوف يصوتون لأن ذلك ما يرغب القائمون على الاستطلاع في سماعه. فقد وجدت الدراسات التي قارنت بين السلوك التصويتي المبلَّغ عنه ذاتيًّا وبين السجلات الانتخابية، أنه ما بين ربع المتجاوبين مع استطلاعات الرأي إلى ثلثهم دائمًا ما يقولون إنهم قد صوتوا رغم أنهم لم يفعلوا ذلك[٤]. من الطرق المستخدمة في تقليص هذا التحيز المحتمل هو أن يُسأل المتجاوب عما إذا كان قد صوَّت في الانتخابات السابقة أو في آخر عدة انتخابات جرت. فمن صوتوا في عدة انتخابات سبقت هم، على الأرجح، سيصوتون في الانتخابات القادمة. وبالمثل فإذا برزت مخاوف لأن يتردد المتجاوبون في أن يجيبوا إجابات غير مقبولة اجتماعيًّا، على غرار وجهة نظر سلبية في مجموعة عرقية أو إثنية معينة، فهنا يجب أن يصاغ السؤال بشكل حاذق على غرار: إذا كان «أناس تعرفهم» يتبنون هذا الرأي.

واحد من أكثر الاستبيانات حساسية هو دراسة أجريت من قِبل المركز القومي لأبحاث الرأي (NORC) بجامعة شيكاجو بعنوان «التنظيم الاجتماعي للنشاط الجنسي: الممارسات الجنسية في الولايات المتحدة» «The Social Organization of Sexuality: Sexual Practices in the United States». وهي الدراسة التي سرعان ما اشتهرت باسم «دراسة الجنس»[٥]. تضمن الوصف الرسمي لهذه الدراسة عبارات على غرار «تنظيم

السلوكيات المكونة للعمليات الجنسية»، و«الاقتران والسلوك الجنسي على الدورة الحياتية». (لست حتى متأكدًا من معنى «الدورة الحياتية»، ويقول برنامج التحقق من الإملاء إنها ليست كلمة حقيقية). سأكون مغاليًا في التبسيط عندما أقول إن هذا الاستبيان نَشَدَ توثيق ماذا يفعل مَن لِمن وبأي وتيرة يفعله. كانت تلك الدراسة التي نشرت عام ١٩٩٥ لا تهدف فقط إلى مدِّنا بمعلومات عن السلوك الجنسي الذي يسلكه جيراننا (رغم أن ذلك كان جزءًا منه)، لكن هدفت أيضًا إلى قياس كم يؤثر السلوك الجنسي في الولايات المتحدة على انتشار فيروس الإتش آي في/ الإيدز.

ولما كان الأمريكيون يترددون في الاعتراف بأنهم لم يصوتوا في الانتخابات، فلك أن تتخيل كيف سيكون حماسهم لوصف سلوكياتهم الجنسية، خاصة إذا ما تضمنت هذه السلوكيات أنشطة غير مشروعة أو خيانة أو حتى أمورًا غير معتادة. لقد كانت منهجية دراسة الجنس مذهلة؛ فقد قامت الدراسة على مقابلات مدةُ كلٍّ منها تسعون دقيقة جرت مع ٣٣٤٢ بالغًا اختيروا ليكونوا ممثلين لمجتمع الأمريكيين البالغين. استكمل ثمانون بالمائة تقريبًا من المتجاوبين المختارين ذلك الاستبيان، مما جعل الباحثين يستخلصون أن نتائجه تعتبر تبيانًا دقيقًا لسلوك أمريكا الجنسي (أو على الأقل ما كنا نفعله عام ١٩٩٥).

ولأنك قد عانيت خلال هذا الفصل من منهجية استطلاعات الرأي، كما أنك قاربت على الانتهاء من كتاب كامل يتناول الإحصاء، فلك الحق الآن في أن تلقي نظرة على النتائج التي وجدوها (وهي نتائج ليس من بينها أي شيء صادم). فكما علق أحدُ مَن راجعوا تلك النتائج قائلًا: «ليس هناك من سلوك جنسي جارٍ بالدرجة التي نظنها»[٦]

- بشكل عام يمارس الناس الجنس مع من يشابهونهم. فتسعون بالمائة من الأزواج يشتركون في نفس العرق والدين والطبقة الاجتماعية والفئة العمرية العامة.
- معظم المتجاوبين مع الاستبيان يمارسون النشاط الجنسي «مرات قليلة في الشهر الواحد»، رغم أنه كانت هناك تغايُرات كبيرة. وتراوَح عدد الشركاء الجنسيين ابتداءً من عمر الثامنة عشرة بين صفر و١٠٠٠
- ٥ بالمائة تقريبًا من الرجال و٤ بالمائة تقريبًا من النساء ذكروا أنهم قد مارسوا نشاطًا جنسيًّا مع شريك من نفس الجنس.
- ثمانون بالمائة من المتجاوبين مع الاستبيان كان لهم شريك جنسي واحد خلال العام الفائت، أو لم يكن لهم شريك أساسًا.
- المتجاوبون ذوو الشريك الجنسي الواحد كانوا أكثر سعادة ممن لم يكن لهم شريك على الإطلاق أو أولئك ذوي الشركاء المتعددين[٧]
- ربع الرجال المتزوجين و١٠ بالمائة من النساء المتزوجات أقروا بأنهم مارسوا

نشاطًا جنسيًا خارج إطار الزواج.

- معظم الناس كانوا يمارسون الجنس بالطريقة القديمة المألوفة: فقد كان الجماع المهبلي أكثر الأنشطة الجنسية قبولًا لدى الرجال والنساء.

أشارت إحدى المراجعات التي تمت على دراسة الجنس إلى انتقاد بسيط، رغم قوته، يتمثل في أن استنتاج كون دقة ذلك الاستبيان تمثِّل الممارسات الجنسية للبالغين في الولايات المتحدة الأمريكية «تفترض أن المتجاوبين مع دراسة المركز القومي لأبحاث الرأي قد عكسوا رأي المجتمع الذي اختيروا منه، وكذلك قدموا إجابات دقيقة»[^٨]. وهذه الجملة يمكن أن تكون الحقيقة التي خرجنا بها من هذا الفصل بأكمله. فللوهلة الأولى قد يبدو الأمر الأكثر إثارة للريبة في عملية استطلاع الرأي هو أن آراء قلة قليلة من الناس يمكن أن تخبرنا بآراء كثيرين غيرهم. لكن ذلك هو الجزء الأيسر. فواحد من أكثر المبادئ الإحصائية أساسيةً يتلخص في أن العينة المنتقاة جيدًا ستأتي مشابِهة للمجتمع الذي انتُقيت منه. أما التحدي الحقيقي في عملية استطلاع الآراء فينقسم إلى قسمين: أولًا العثور على العينة المناسبة ثم الوصول إليها، وثانيًا استخراج المعلومات من تلك المجموعة الممثلة بطريقة تعكس بدقة ما يعتقده أفرادها.

ملحق الفصل العاشر

لماذا يصير الخطأ المعياري كبيرًا عندما يكون P و(1-P) قريبين من ٥٠ بالمائة؟

إليك المنطق الذي يقف وراء كون الخطأ المعياري يكون في أعلى حالاته عندما يجيب أفراد العينة بطريقة معينة تكون فيها (p) قريبة من ٥٠ بالمائة (وهو الأمر الذي يعني من خلال الحسابات الرياضية البسيطة أن 1-p ستكون أيضًا قريبة من ٥٠ بالمائة). لنتخيل أنك تجري استطلاعَي رأي في نورث داكوتا. أول هذين الاستطلاعين مصمم لقياس نسبة الجمهوريين إلى نسبة الديمقراطيين في الولاية. لنفترض أن النسبتين الحقيقيتين للمعسكرين في الولاية منقسمتين النصف للنصف، لكن استطلاع الرأي الذي أجريته خرج بنسبة ٦٠ بالمائة للجمهوريين في مقابل ٤٠ بالمائة للديمقراطيين. نتائجك تلك حادت عن النسبة الصحيحة بنسبة ١٠ بالمائة وهو هامش كبير. لكنك مع ذلك قد تسببت في هذا الخطأ الكبير دون أن ترتكب خطأً كبيرًا جدًا أثناء جمعك للبيانات. فقد بالغت في تعداد الجمهوريين متعديًا نسبتهم الحقيقية بـ ٢٠ بالمائة [(٦٠ – ٥٠)/ ٥٠]. وبهذا قد قمت بالتقليل من العدد الحقيقي للدمقراطيين بـ ٢٠ بالمائة [(٤٠ – ٥٠)/ ٥٠]. وذلك ممكن الحدوث حتى مع اتباع منهجية سليمة في استطلاع الآراء.

أما استطلاع الرأي الثاني فقد صُمم لتقدير عدد الأمريكيين الأصليين في نورث داكوتا. لنفترض أن الأمريكيين الأصليين في نورث داكوتا يمثلون ١٠ بالمائة من تعداد السكان، في حين يمثل غيرهم ٩٠ بالمائة من تعداد الولاية. الآن لنناقش كيف هو السوء الذي يجب أن تكون عليه عملية جمع البيانات كي ينتج عنها استطلاع رأي ذو خطأ استعيان قدره ١٠ نقاط مئوية. يمكن أن يحدث ذلك بطريقتين: أولاهما أنك يمكن أن تجد بالمائة من السكان ينتمون إلى الأمريكيين الأصليين و ١٠٠ بالمائة من غيرهم. أو قد تجد ٢٠ بالمائة من السكان أمريكيين أصليين و ٨٠ بالمائة من غيرهم. في الحالة الأولى قد أغفلتَ جميع الأمريكيين الأصليين، وفي الحالة الثانية قد وصلت إلى ضعف العدد الحقيقي لهم. تلك هي فعلًا أخطاء استعيان سيئة. وفي كلتا الحالتين أخطأ تقديرك بنسبة ١٠٠ بالمائة سواء [(٠ – ١٠)/١٠] أو [(٢٠ – ١٠)/١٠]. لو كنت قد أخطأت في نسبة الأمريكيين الأصليين خطأً قدره ٢٠ بالمائة فقط – نفس نسبة الخطأ في استطلاع الجمهوريين والديمقراطيين – لجاءت نتائجك النهائية في صورة ٨ بالمائة للأمريكيين الأصليين و٩٢ بالمائة لغيرهم، وهو ما يعني فارق نقطتين مئويتين فقط عن النسبة الحقيقية من السكان.

عندما يقترب p و١– p من ٥٠ بالمائة، تتضخم أخطاء الاستعيان الصغيرة فتصير أخطاءً كبيرة جدًّا في مخرج استطلاع الرأي.

عندما يقترب p أو 1-p من الصفر، يحدث العكس، فتتضاءل حتى أخطاء الاستعيان الكبيرة نسبيًّا وتصير أخطاءً صغيرة جدًّا في مخرج استطلاع الرأي.

نفس خطأ الاستعيان الذي قدره ٢٠ بالمائة قد شوَّه مخرج استطلاع رأي الجمهوريين والديمقراطيين بمقدار ١٠ نقاط مئوية، في حين أنه شوَّه استطلاع رأي الأمريكيين الأصليين بمقدار نقطتين مئويتين. وبما أن الخطأ المعياري يقاس في استطلاع الرأي بقيم مطلقة (مثل ± ٥ بالمائة)، تنص المعادلة على أن الخطأ سيكون أكبر على الأرجح عندما يكون كلٌّ من p و1-p قريبَين من ٥٠ بالمائة.

الفصل الحادي عشر

تحليل الانحدار

الإكسير المعجزة

هل يمكن أن يقتلك توتر العمل؟ نعم. هنالك دليل دامغ على أن مَشاقَّ العمل قد تؤدي إلى الموت المبكر، وخاصة من جراء مرض القلب. لكنه ليس نوع التوتر الذي قد يكون دار بمخيلتك. فالمديرون التنفيذيون الذين لا يفتئون يتخذون قرارات مهمة جدًّا تحدد مصائر شركاتهم، هم أقل عرضة بكثير للموت المبكر من مساعديهم الذين تقتضي مهام عملهم الرد على الاتصالات الهاتفية وغيرها من المهام التي يُلَقَّنون عملها. كيف يمكن أن يكون ذلك منطقيًّا؟ لقد اتضح أن أشد أنواع التوتر خطورة ينبع من أن يكون للمرء «مستوًى ضعيف من التحكم» في مسئولياته. وَجدت الكثير من الدراسات التي أجريت على آلاف البريطانيين العاملين بالخدمة المدنية (دراسات وايت هول) أن العاملين الذين لا يحوزون إلا القليل من التحكم في وظائفهم – بمعنى أن سلطتهم محدودة جدًّا على المهام التي يؤدونها أو كيفية تأديتها – لهم معدلات وفيات أعلى بكثير من العاملين الآخرين ذوي سلطة اتخاذ القرارت. تقول تلك الدراسة إن التوتر الناتج عن المسئوليات الجسام ليس هو الذي سيقتلك، وإنما التوتر المصاحب لأن يملى عليك ما تفعل دون أخْذ رأيك فيما يُفعل أو كيفية فعله هو الذي سيقتلك.

ليس هذا بفصل يتناول توتر العمل أو مرض القلب أو العاملين المدنيين البريطانيين. السؤال المهم هنا الذي يتعلق بدراسات وايت هول (والدراسات الأخرى التي تشبهها)، هو كيف يمكن للباحثين أن يَخلُصوا إلى تلك النتيجة؟ من الواضح أنه لا يمكن أن تكون تلك تجربة أجريت بعشوائية. فلا يمكن أن نعيِّن بشرًا بشكل اعتباطي في وظائف مختلفة ونجبرهم على أن يعملوا في هذه الوظائف لسنين عديدة، ثم نرى من منهم يموت بمعدلات أكبر. (لو نحَّيْنا الاعتبارات الأخلاقية جانبًا، لوجدنا أن توزيع الوظائف بشكل عشوائي من شأنه أن يشيع الفوضى في مجال الخدمة المدنية في بريطانيا). لذلك فقد جمع الباحثون بيانات طولية عن آلاف الأفراد العاملين بالخدمة المدنية في بريطانيا، وهذه البيانات يمكن

تحليلها لتحديد الارتباطات ذات المغزى، كما في الارتباط بين الوظائف «ذات مستوى التحكم الضعيف» ومرض شرايين القلب التاجية.

لكن ارتباطًا بسيطًا كهذا لا يكفي لاستنتاج كون بعض أنواع الوظائف ضارًّا بالصحة. فإذا اكتفينا بملاحظة أن العاملين ذوي الرتب الدنيا في سُلَّم وظائف الخدمة المدنية البريطانية هم ذوو معدلات عالية للإصابة بأمراض القلب، فسنجد نتائجنا تشوِّشها عوامل أخرى. فمثلًا سنجد العاملين ذوي المستويات الدنيا لم يحظوا بتعليم عالٍ بنفس مستوى كبار المسئولين في النظام. وكذلك هم أكثر ميلًا للتدخين (ربما بسبب ما يلاقون من إحباطات في عملهم)، وقد يكونون قد مروا بطفولة أقل صحية، وهو الأمر الذي قلص من تطلعاتهم الوظيفية. أو أن دخولهم الضعيفة قد حدَّت من تمتعهم برعاية صحية جيدة. وهلم جرًّا. الفكرة هنا أن أي دراسة لا تفعل أكثر من المقارنة بين المخرجات الصحية لمجموعة كبيرة من العاملين البريطانيين – أو في داخل أي مجموعة أخرى كبيرة – لن تخبرنا بالكثير. فمصادر البيانات الأخرى سوف تلبس علينا العلاقة التي تعنينا. هل يسبب «التحكم الضعيف في الوظيفة» مرض القلب فعلًا؟ أم إن السبب يكمن في تجمُّع من العوامل الأخرى التي تَصادَف أنها مشتركة لدى من يحوزون مستوًى أقل من التحكم في وظائفهم؟ وفي هذه الحالة قد نغفل تمامًا عن التهديد الحقيقي للصحة العامة.

تحليل الانحدار هو الأداة الإحصائية التي تساعدنا على مواجهة هذا التحدي. وبشكل أكثر تحديدًا يسمح لنا تحليل الانحدار بتقدير العلاقة بين متغير معين ومخرج نُعنى به، بينما نسيطر على عوامل أخرى. بعبارة أخرى نقول إننا نستطيع أن نعزل تأثير متغير واحد، مثل تولي نوع معين من الوظائف، بينما نتمسك بمتغيرات أخرى كثابت. استخدمت دراسات وايت هول تحليل الانحدار لتقدير التأثيرات الصحية للوظائف ذات التحكم الضعيف بين من يتشابهون في أمور أخرى كسلوك التدخين. (العاملون ذوو الرتب الدنيا يدخنون بالفعل أكثر من رؤسائهم، وهو ما يفسر كمية صغيرة نسبيًّا من المتغيرات في مرض القلب في التراتب الهرمي في دراسات وايت هول).

تستند غالبية الدراسات التي تقرأ عنها في الصحف إلى تحليل الانحدار. فعندما يخلص الباحثون إلى أن الأطفال الذين يقضون كثيرًا من الوقت في دور الرعاية النهارية هم أكثر عرضة للمشكلات السلوكية في المدارس الابتدائية من أقرانهم الذين قضوا جُلَّ وقتهم في منازلهم، فلا يعني هذا أنهم قد وزعوا آلاف الرضع بين دور الرعاية النهارية وبين المنازل مع أحد الوالدين، ولا اكتفى الباحثون بمقارنة سلوكيات أطفال المدارس الابتدائية ذوي تجارب الطفولة المتباينة دونما اعتبار لكون هذين المجتمعين يتباينان على الأرجح بطرائق أساسية أخرى. فالأسر المختلفة تتخذ قرارات مختلفة فيما يتعلق برعاية الأطفال لأنهم مختلفون. فبعض الأسر بها أبوان حاضران وبعضها ليست كذلك. وفي بعضها يعمل

الوالدان وفي البعض الآخر ليس الأمر كذلك. وبعض الأسر أكثر ثراءً وأعلى مستوًى علميًّا من البعض الآخر. كل هذه الأمور تؤثر على قرارات رعاية الأطفال وتؤثر على أداء أطفال هذه الأسر في المدارس الابتدائية. إذا أجري تحليل الانحدار بالشكل السليم، فسيساعدنا على تقدير آثار دور الرعاية النهارية منفصلةً عن الأمور الأخرى التي تؤثر على الأطفال، كدخل الأسرة وهيكل الأسرة ومستوى تعليم الوالدين، وهلم جرًّا.

في الجملة السابقة وردت عبارتان أساسيتان. أولاهما هي: «إذا أجري تحليل الانحدار بالشكل السليم». إذا ما توافرت البيانات السليمة وتوافر حاسوب شخصي، يستطيع ابن ست سنوات أن يستخدم برنامج إحصاء بدائيًّا فيخرج بنتائج الانحدار. فقد جَعلت الحواسيبُ الشخصية آلياتِ تحليل الانحدار غايةً في اليسر. فليست آليات تحليل الانحدار الجزء الصعب في المسألة، وإنما تكمن الصعوبة في تحديد أيُّ المتغيرات ينبغي أن يضعها التحليل في اعتباره، وكيف هي الطريقة المثلى لإنجاز هذا التحليل. وتحليل الانحدار يشبه واحدة من تلك الآلات الكهربائية الفاخرة التي هي سهلة الاستخدام نسبيًّا، لكن من العسير استخدامها بالشكل الأمثل – بل إنها خطيرة إذا ما استخدمت بالشكل الخطأ.

أما العبارة المهمة الثانية، فهي «سيساعدنا على تقدير». فدراسة الرعاية النهارية للأطفال تلك التي بين أيدينا لا تمنحنا الإجابة «الصحيحة» فيما يتعلق بالعلاقة بين دور الرعاية النهارية والأداء اللاحق في المدرسة الابتدائية. وإنما تقدِّر تقديرًا كميًّا العلاقة التي تمَّت ملاحظتها لدى مجموعة من الأطفال على مدى فترة معينة من الزمن. هل نستطيع أن نستخلص نتائج قد تنطبق على المجتمع الأوسع؟ نعم، لكن سنواجه نفس المحددات والشروط التي تواجهنا مع أي نوع من أنواع الاستدلال. أولًا لا بد أن تكون عيِّنتنا ممثلة للمجتمع الذي نُعنى به. فلو أجريت دراسة على ٢٠٠٠ طفل صغير في السويد لن تخبرنا بالكثير عن أفضل سياسات تعليم الأطفال في المناطق الرعوية بالمكسيك. ثانيًا سوف يكون هناك تغاير بين عينة وأخرى. فلو أجرينا دراسات متعددة عن الأطفال ورعاية الطفولة، فسنخرج من كل دراسة بنتائج ذات تباينات طفيفة، حتى لو كانت المنهجيات سليمة ومتشابهة.

يتشابه تحليل الانحدار مع إجراء استطلاعات الرأي. لكن الخبر الجيد هنا هو أننا لو حُزْنَا عينة كبيرة ممثلة ومنهجية متماسكة، فلن يرجح أن تنحرف العلاقة التي نلاحظها من خلال بيانات العينة انحرافًا شديدًا عن العلاقة الحقيقية القائمة في المجتمع كاملًا. إذا وجدنا أن لألف شخص يمارسون الرياضة ثلاث مرات أو أكثر أسبوعيًّا معدلاتٍ للإصابة بأمراض الشرايين القلبية تتدنى كثيرًا عن مثيلتها لدى ألف آخرين لا يمارسون الرياضة (مع تشابههم في جميع الاعتبارات المهمة الأخرى)، هنا تتزايد احتمالات وجود ارتباط مشابه بين ممارسة الرياضة وصحة الشرايين القلبية في المجتمع الأوسع؛ لذلك نجري تلك

الدراسات. (وليس الهدف هنا هو أن نقول لهؤلاء المرضى الذين لا يمارسون الرياضة إنهم كان يجدر بهم أن يمارسوا الرياضة في الماضي).

أما الخبر السيء فهو أننا لا نقطع بأن ممارسة الرياضة يقي من مرض القلب. وإنما ما نفعل هو أننا نرفض الفرضية المنعدمة، والتي تنص على أن ممارسة الرياضة ليست ذات ارتباط بمرض القلب، استنادًا إلى حد إحصائي معين تم اختياره قبل إجراء الدراسة. بشكل أكثر تحديدًا نقول إن القائمين على هذه الدراسة سوف يقولون إنه لو كانت ممارسة الرياضة ليست ذات صلة بصحة الشرايين القلبية، فإن أرجحية ملاحظة فارق معتبر في الإصابة بأمراض القلب بين ممارسي الرياضة ومن لا يمارسونها في هذه العينة الكبيرة سيكون أقل من ٥ من كل ١٠٠، أو ما دون أي حد آخر للدلالة الإحصائية.

لنتوقف للحظة ونلوِّح بأول أعلامنا الصفراء الكبيرة. لنفترض أن هذه الدراسة بالذات كانت تقارن بين مجموعة كبيرة ممن يمارسون رياضة الاسكواش بانتظام، ومجموعة أخرى مساوية لها لا يمارس أفرادها اللعبة مطلقًا. ممارسة الاسكواش ليست تمرينًا مفيدًا للشرايين القلبية. لكننا نعلم أن لاعبي الاسكواش عادةً ما يكونون على درجة من الثراء تسمح لهم بأن يلتحقوا بنوادٍ بها ساحات اسكواش. والأثرياء قادرون على الحصول على الرعاية الصحية، وهو الأمر الذي من شأنه أن يحسِّن من صحة الشرايين القلبية. التحليل السيء قد يدفعنا لأن نعزو فوائد صحية إلى ممارسة الاسكواش، في حين أن تلك الفوائد في الحقيقة ناتجة من كون اللاعبين بالثراء الكافي الذي يجعلهم يمارسون الاسكواش (وفي هذه الحالة سترتبط ممارسة البولو أيضًا بتحسن صحة القلب، رغم أن الحصان في هذه اللعبة هو الذي يؤدي غالبية المجهود).

أو لربما كانت السببية تسير في الاتجاه الآخر. هل يمكن لامتلاك قلب سليم أن «يسبب» ممارسة الرياضة؟ نعم. فمعتلُّو الصحة، وخاصة من يعانون بوادئ مرض القلب منهم، يتعذر عليهم جدًّا أن يمارسوا الرياضة. وبالقطع ستقل احتمالات ممارستهم لرياضة الاسكواش بانتظام. مرة أخرى لو كان التحليل سيئًا أو مبالغًا في تبسيطه، فحينها سيكون الادعاء بأن ممارسة الرياضة يعكس حقيقة أن معتلِّي الصحة يجدون في ممارسة الرياضة مشقة كبيرة. وفي هذه الحالة لن تجعل ممارسة الاسكواش أي أحد أكثر صحة، وإنما لن تفعل سوى أنها ستفرق بين الصحيح والعليل.

هنالك الكثير من الشراك المحدقة في الانحدار، حتى إنني قد خصصت الفصل التالي بأكمله لعرض أكثر الأخطاء المتعلقة به فداحة. أما الآن فسوف نركز على الأمور التي ستجري في مجراها الصحيح. فتحليل الانحدار يمتلك قدرة مدهشة على عزل العلاقة الإحصائية التي تهمنا، كتلك العلاقة التي بين التحكم في العمل ومرض القلب، مع الأخذ في الاعتبار العوامل الأخرى التي قد تشوِّش تلك العلاقة.

كيف يحدث ذلك تحديدًا؟ لو علمنا أن موظفي الخدمة المدنية البريطانيين ذوي الرتب الدنيا يدخنون أكثر مما يفعل رؤساؤهم، فكيف لنا أن نحدد أيُّ جزء من اعتلال شرايينهم القلبية يعزى إلى تدنِّي رتبهم الوظيفية وأيها مردُّه إلى التدخين؟ فهذان العاملان يبدوان متداخلين بشكل معقد.

تحليل الانحدار (إذا ما أنجز بشكل سليم) قادر على فك تعقيدهما. لشرح المنطق من وراء هذا، أحتاج لأن أبدأ بالفكرة الأساسية التي تستند إليها كل صور تحليل الانحدار – من أبسط العلائق الإحصائية إلى النماذج المعقدة التي يؤلفها علماء حائزون على جوائز نوبل. فتحليل الانحدار في جوهره ينشد إيجاد «أفضل موضع» تقع فيه العلاقة الخطية بين متغيرين. من الأمثلة البسيطة على ذلك العلاقةُ بين الطول والوزن. فطِوال القامة عادة ما يكونون أثقل وزنًا – رغم أن ذلك لا يصح دائمًا. فإذا ما وزعنا أطوال مجموعة من طلبة الدراسات العليا وأوزانهم توزيع انتشار، فلعلك ستذكر كيف بدا ذلك التوزيع عندما أوردناه في الفصل الرابع:

توزيع انتشار للطول والوزن

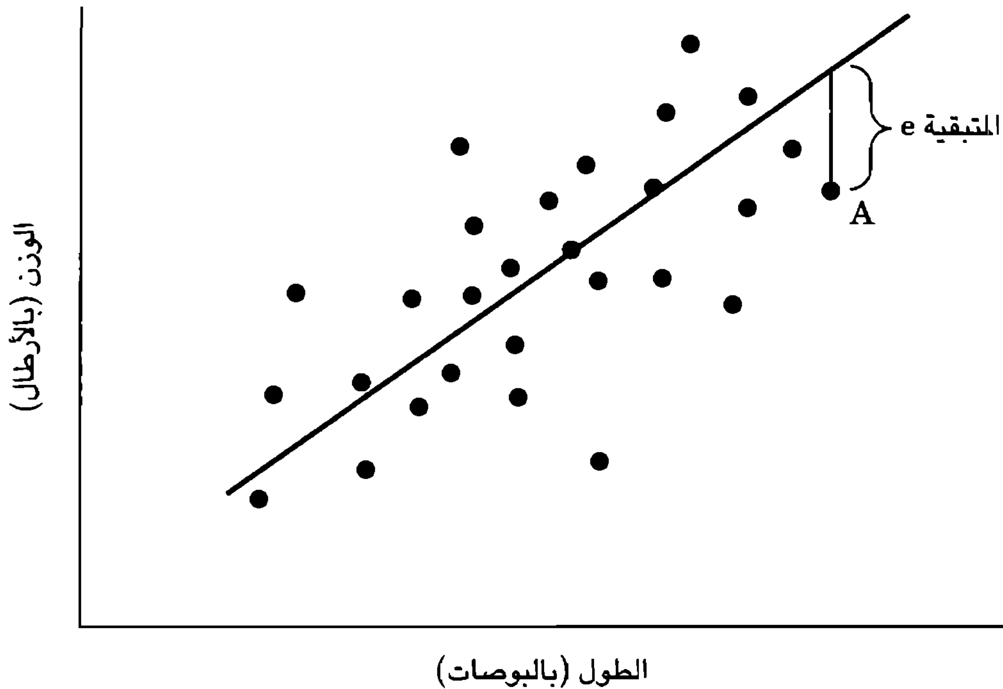


إذا ما طُلب منك أن تصف النمط في هذا الشكل، فقد تقول شيئًا على غرار «يبدو كما لو أن الوزن يزيد مع زيادة الطول». وتلك ليست بالمقولة المفيدة جدًّا أو المحددة. لكن تحليل الانحدار يمكننا من أن نمضي خطوة إضافية و«نمد خطًّا عبر التوزيع» هو أحسن ما يصف العلاقة الخطية القائمة بين متغيرين.

هنالك خطوط كثيرة تتماشى نوعًا ما مع بيانات الطول والوزن. لكن كيف لنا أن نعرف أيُّ الخطوط الأفضل لهذه البيانات؟ بل كيف لنا أن نُعرِّف كلمة «الأفضل» بالضبط؟ عادة ما يَستخدم تحليل الانحدار منهجية تسمى المربعات الصغرى العادية أو OLS. لكن التفاصيل التقنية التي تفسر أمورًا، من بينها سبب كون المربعات الصغرى العادية تُنتج أفضل الخطوط، فلا بد أن نتركها لكتَّاب أكثر تقدمًا. تكمن النقطة الرئيسية في ذلك الجزء من اسمها «المربعات الصغرى»، فهي تتماشى مع الخط الذي يقلل مجموع المتبقيات المربعة. ليس الأمر بالتعقيد الشديد الذي يبدو عليه. فكل مشاهدة في بيانات الطول والوزن التي لدينا لها متبقيات، والتي هي المسافة الرأسية من خط الانحدار، عدا تلك المشاهدات التي تقع مباشرة على الخط، والتي تساوي فيها المتبقيات صفرًا. (في الشكل الوارد بالأسفل تم تعيين المتبقية لشخص افتراضي A). ومن المنطقي أنه كلما زاد المجموع الكلي للمتبقيات، ساء موقع ذلك الخط. لكن المنعطف الوحيد الذي ينافي المنطق في المربعات الصغرى العادية، هو أن المعادلة تأخذ مربع كل متبقية قبل أن تجمعهم كلهم (وهو ما يزيد من ثقل المشاهدات التي تقع بعيدًا عن خط الانحدار أو «القيم المتطرفة»).

المربعات الصغرى العادية «تصل» الخط الذي يقلص من مجموع المتبقيات المربعة، كما هو موضح في الشكل التالي:

الخط الأنسب للطول والوزن



لو أصابتك التفاصيل التقنية بالصداع، فيكفيك فقط أن تفهم الناتج النهائي

الذي ينص على أن المربعات الصغرى العادية تعطينا أفضل وصف للعلاقة الخطية بين متغيرين. والنتيجة لا تقتصر على خط، وإنما كذلك – كما تذكر من مادة الهندسة في المدرسة الثانوية – تحتوي على معادلة تصف ذلك الخط. تُعرف تلك المعادلة باسم معادلة الانحدار، وتتخذ الشكل التالي: $y = a + bx$؛ حيث y هو الوزن بالأرطال وa هو تداخل y مع الخط (قيمة y عندما يكون x = ٠)، وb هو ميل الخط وx هو الطول بالبوصات. ميل الخط الذي وضعناه، b، يصف «أفضل» علاقة خطية بين الطول والوزن لهذه العينة كما تعرفها المربعات العادية الصغرى.

خط الانحدار لا يصف بكل تأكيد كل مشاهدة في مجموعة البيانات بشكل مثالي. لكنه أفضل وصف يمكن أن نخرج به لتلك العلاقة الواضحة ذات المدلول والقائمة بين الطول والوزن. وهو كذلك يعني أنه يمكن أن تفسر كل مشاهدة على النحو التالي: الوزن = a + b (الطول) + e، حيث e هي «متبقية» تلتقط التغاير في الوزن الذي لا يفسره الطول لدى كل فرد من أفراد العينة. وأخيرًا فهو يعني أن أفضل تقديراتنا لوزن أي شخص في مجموعة البيانات سوف يكون a + (الطول) b + e. ورغم أن غالبية المشاهدات لا تقع بالضبط على خط الانحدار، لكن المتبقية لا تزال قيمتها المتوقعة صفرًا، بما أن أي شخص ضمن عينتنا يرجح أن يزيد وزنه أو ينقص على ما تقول به معادلة الانحدار.

كفانا من هذه الرطانة النظرية، ولنفحص بعض البيانات الحقيقية من دراسة الحيوات المتغيرة، لكنني أحتاج أولًا لأنْ أوضِّح بعض المصطلحات الأساسية. المتغير الذي يتم تفسيره – الوزن في هذه الحالة – يعرف باسم المتغير الاعتمادي (لأنه يعتمد على عوامل أخرى). والمتغيرات التي نستخدمها لتفسير المتغير الاعتمادي تعرف باسم المتغيرات المفسرة لأنها تفسر الناتج الذي نهتم به. (ولجعل الأمر أكثر صعوبة، أحيانًا تسمى المتغيرات المفسرة بالمتغيرات المستقلة أو متغيرات التحكم). لنبدأ باستخدام الطول في تفسير الوزن لدى المشاركين في دراسة الحيوات المتغيرة، وفيما بعد سوف أضيف بعض العوامل الأخرى التي قد تكون مفسرة أيضًا*. هناك ٣٥٣٧ مشاركًا بالغًا في دراسة الحيوات المتغيرة. ذلك هو عدد المشاهدات أو n (أحيانًا تشير الورقة البحثية إلى أن n = ٣٥٣٧). عندما نجري تحليل انحدار بسيط على بيانات دراسة الحيوات المتغيرة، واضعين الوزن متغيرًا اعتماديًّا والطول هو المتغير المفسر الوحيد، فسنحصل على النتائج التالية:

الوزن = – ١٣٥ + (٤,٥) × الطول بالبوصات.

a = – ١٣٥. ذلك هو تداخل الـ – y، والذي لا يحمل أي مدلول في حد ذاته. (لو ترجمنا ما سبق بشكل حرفي لكان الشخص الذي يبلغ طوله صفر بوصة يزن سالب

* يجدر بك أن ترى في هذا التمرين «مرحًا مع البيانات» لا استكشافًا رسميًّا لأي من العلاقات التي ستوصف في معادلات الانحدار اللاحقة. الهدف هنا هو تقديم مثال منطقي لكيفية عمل تحليل الانحدار، لا إجراء بحث ذي دلالة عن الأوزان في أمريكا.

١٣٥ رطلًا، وبالطبع فهذا هراء على مستويات عدة). يعرف هذا الرقم كذلك باسم الثابت لأنه نقطة البداية لحساب أوزان جميع مشاهدات الدراسة.

b = ٤,٥. تقديرنا لـ b، والبالغ ٤,٥، يعرف باسم معامل الانحدار، أو كما تسميه رطانة الإحصاء «معاملًا للطول»؛ وذلك لأنه يمنحنا أفضل تقدير للعلاقة القائمة بين الطول والوزن لدى المشاركين في دراسة الحيوات المتغيرة. ولمعامل الانحدار تفسير مناسب، ألا وهو أن زيادة وحدة واحدة في المتغير المستقل (الطول) ترتبط بزيادة في ٤,٥ وحدات من المتغير الاعتمادي (الوزن). يعني هذا في بيانات عينتنا أن زيادة قدرها بوصة واحدة في الطول ترتبط بزيادة قدرها ٤,٥ أرطال في الوزن. لذلك فإذا لم تتوافر لدينا أي معلومات أخرى، فسيكون أفضل تخميناتنا لوزن أحد المشاركين في دراسة الحيوات المتغيرة طوله ٥ أقدام و١٠ بوصات (٧٠ بوصة) يساوي – ١٣٥ + ٤,٥ (٧٠) = ١٨٠ رطلًا.

تلك هي النتيجة الحاسمة، فقد قمنا بتقدير كمي لأفضل وضع للعلاقة الخطية بين الطول والوزن لدى المشاركين في دراسة الحيوات المتغيرة. نفس تلك الأدوات الأساسية يمكن أن تستخدم لاستكشاف علائق أكثر تعقيدًا وأسئلة أكثر دلالة من الناحية الاجتماعية. ففي كل معامل انحدار سوف تكون مهتمًّا بشكل عام بثلاثة أشياء، ألا وهي الإشارة والحجم والدلالة.

الإشارة: الإشارة (موجبة كانت أو سالبة) التي تَسِمُ معامل المتغير المستقل تخبرنا باتجاه الارتباط مع المتغير الاعتمادي (المخرج الذي نحاول تفسيره). في الحالة البسيطة السابقة كان المعامل للطول موجبًا. فكان طِوال القامة أميَل للزيادة في الوزن. لكن بعض العلائق تمضي في الاتجاه الآخر. فإنني أتوقع أن يكون الارتباط بين ممارسة الرياضة وزيادة الوزن سالبًا. لو تضمنت دراسة الحيوات المتغيرة بيانات عن شيء على غرار «عدد الأميال التي يجريها الفرد في الشهر»، فإنني على شبه يقين أن المعامل «للأميال التي جُرِيَت» سيكون سالبًا. فالمزيد من الجري هو أمر يرتبط بقلة الوزن.

الحجم: كم قدر التأثير الذي نلاحظه بين المتغير المستقل والمتغير الاعتمادي؟ هل هو كم معتبر؟ في هذه الحالة ترتبط كل بوصة واحدة بـ ٤,٥ أرطال، والتي هي نسبة مئوية معتبرة في وزن جسد شخص عادي. وهكذا يكون الطول عاملًا مهمًّا من عوامل تفسير لماذا يزن بعض الأشخاص أكثر مما يزن غيرهم. في دراسات أخرى قد نجد ذلك المتغير المفسر ذا التأثير الإحصائي المهم على المخرج الذي نحن معنيون به – بمعنى أن ذلك التأثير الملاحظ لا يرجح حدوثه عن طريق الصدفة – لكن ذلك التأثير قد يكون ضعيفًا جدًّا وتافهًا جدًّا أو غير ذي دلالة اجتماعيًّا. فمثلًا لنفترض أننا نفحص محددات الدخل. لماذا يجني بعض الناس أموالًا أكثر من غيرهم؟ هنا يرجح أن تكون المتغيرات المفسرة أشياء على غرار التعليم وسنوات الخبرة العملية، وهلم جرًّا. وقد يجد الباحثون أيضًا في مجموعة

كبيرة من البيانات أن من لهم أسنان أكثر بياضًا يجنون ٨٦ دولارًا سنويًّا أكثر من غيرهم من العاملين إذا ما تساوت بقية العوامل. يفترض المعامل الإيجابي والدال إحصائيًّا المتعلق بـ «الأسنان البيضاء» أن الأشخاص الذين تقع المقارنة عليهم متشابهون في اعتبارات أخرى، كنفس مستوى التعليم ونفس سنوات الخبرة العملية، وهلم جرًّا. (سوف أوضح بعد برهة كيف لنا أن ننجز ذلك العمل المضني). لقد أوضح تحليلنا الإحصائي أن الأسنان الأكثر بياضًا ترتبط بزيادة في الدخل السنوي قدرها ٨٦ دولارًا، وأن هذه النتيجة ليست محض صدفة. وهو ما يعني أننا (١) قد رفضنا الفرضية المنعدمة، والتي تنص على أن الأسنان البيضاء ليست ذات صلة بالدخل بمستوًى عالٍ من الثقة؛ و(٢) لو حللنا عينات البيانات الأخرى، فمن المرجح أن نجد علاقة مشابهة بين الأسنان الجميلة وارتفاع الدخل.

لكن ما معنى هذا؟ نعم لقد وجدنا نتائج ذات دلالة إحصائية، لكن لم نجد بينها واحدة تحمل معنًى في حد ذاتها. فبادئ ذي بدء ليست الستة والثمانون دولارًا بالمبلغ الذي قد يغير حياة المرء. ومن وجهة نظر السياسة العامة، يحتمل أن تكون الستة والثمانون دولارًا أقل من تكلفة تبييض أسنان شخص واحد كل عام؛ لذا فلا نستطيع حتى أن نوصي بأن يقوم صغار العاملين بهذا الاستثمار. ورغم أنني سأستبق فصلًا كاملًا فإنني سأعرب عن قلقي بشأن بعض المشكلات المنهجية الخطيرة. فعلى سبيل المثال قد تكون الأسنان الجيدة ترتبط بخصال شخصية أخرى تفسر مزايا الدخل المالي الزائد، فقد يكون تأثير زيادة الدخل قد سببته نوعية الأشخاص الذين يُعْنون بأسنانهم لا الأسنان في حد ذاتها. لكن لنكتفِ الآن بتسجيل حجم الارتباط الذي نلاحظه بين المتغير المفسر والمخرج الذي يهمنا.

الدلالة: هل تلك النتيجة التي لاحظناها هي انحراف يقوم على عينة بيانات منحرفة، أم إنها تعكس ارتباطًا ذا معنًى يرجَّح أن يصدُق على المجتمع ككل؟ ذلك هو نفس السؤال الأساسي الذي ظللنا نطرحه على مدى الفصول السابقة. في سياق الطول والوزن، هل نظن أننا سنجد ارتباطًا إيجابيًّا مشابهًا في العينات الممثلة للمجتمع؟ للإجابة عن هذا السؤال نستخدم أدوات استدلال قد عرفناها بالفعل. فمعامل الانحدار الذي لدينا يقوم على علاقة ملحوظة بين الطول والوزن في عينة بيانات بعينها. لكن لو اختبرنا عينة بيانات أخرى كبيرة، فلنا أن نكون على شبه يقين بأننا سوف نجد ارتباطًا يختلف اختلافًا طفيفًا بين الطول والوزن، ومن ثم معاملًا مختلفًا كذلك. فالعلاقة بين الطول والوزن التي تلاحَظ في بيانات وايت هول (عاملي الخدمة المدنية البريطانيين) ستكون مختلفة على الأرجح عن العلاقة الملاحظة بين الطول والوزن في بيانات دراسة الحيوات المتغيرة. لكننا نعلم من مبرهنة النهاية المركزية أن متوسط أي عينة كبيرة منتقاة بعناية لن ينحرف بشدة عن متوسط المجتمع كاملًا. وبالمثل لنا أن نفترض أن العلاقة التي نلاحظها بين متغيرين،

كالطول والوزن، لن تتباين بشدة من عينة لأخرى، مع افتراض أن هذه العينات كبيرة ومنتقاة جيدًا من نفس المجتمع.

فلنفكر في ذلك المنطق: إنه من المستبعد جدًّا (رغم كونه ممكنًا) أن نجد كل بوصة في الطول ترتبط بـ ٤,٥ أرطال إضافية بين المشاركين في دراسة الحيوات المتغيرة، لكن في نفس الوقت ليس هناك ارتباط بين الطول والوزن في عينة أخرى ممثلة قوامها ٣٠٠٠ أمريكي بالغ.

يجدر بهذا أن يمنحك الدليل الأول على كيفية اختبار إذا ما كانت نتائج الانحدار ذات دلالة إحصائية أم لا. وكما في استطلاعات الرأي وغيرها من طرائق الاستدلال، يمكننا أن نحسب الخطأ المعياري لمعامل الانحدار. والخطأ المعياري هو مقياس للتشتت المرجح لعينات مكررة منتقاة من مجتمع واحد. إذا ما كنا بصدد قياس ووزن عينة مختلفة قوامها ٣٠٠٠ أمريكي، فقد نجد في التحليل اللاحق أن كل بوصة في الطول ترتبط بـ ٤,٣ أرطال. ولو كررناها في عينة أخرى مكونة من ٣٠٠٠ أمريكي، فقد نجد كل بوصة ترتبط بـ ٥,٢ أرطال. مرة أخرى نجد التوزيع الطبيعي صديقنا. فيما يتعلق بعينة البيانات الكبيرة التي تخص دراسة الحيوات المتغيرة، لنا أن نفترض أن معاملاتنا المتعددة ستتوزع بشكل طبيعي حول الارتباط «الحقيقي» بين الطول والوزن في مجتمع الأمريكيين البالغين. بناءً على هذا الافتراض يمكننا أن نحسب الخطأ المعياري لمعامل الانحدار، والذي قد يجعلنا نخمن كم من التشتت في المعاملات نتوقع من عينة لأخرى. لن أخوض هنا في معادلة حساب الخطأ المعياري، وذلك لسببين؛ أولهما: أن ذلك سوف يحيد بنا إلى اتجاه يعج بالحسابات الرياضية، والثاني: أن برامج الحزم الإحصائية الأساسية سوف تحمل عنك عناء حسابه بنفسك.

لكن عليَّ أن أحذِّر هنا من أننا إذا عملنا على عينة بيانات صغيرة – كمجموعة مكونة من ٢٠ بالغًا لا ما يزيد عن الثلاثة آلاف شخص في دراسة الحيوات المتغيرة – فحينها لن يعود التوزيع الطبيعي صديقنا. وعلى وجه التحديد إذا ما كررنا إجراء تحليل انحدار على عينات صغيرة مختلفة، لن يعود بإمكاننا افتراض توزُّع معاملاتنا المتعددة طبيعيًّا حول الارتباط «الحقيقي» الحادث بين الطول والوزن في مجتمع البالغين الأمريكيين. وإنما سوف تتوزع معاملاتنا حول الارتباط «الحقيقي» بين الطول والوزن لمجتمع الأمريكيين البالغين فيما يعرف باسم توزيع تي. (توزيع تي هو في الأساس توزيع أكثر تشتتًا من التوزيع الطبيعي، ومن ثم فإن لها «أذيالًا أكثر سُمكًا»). لكن لا شيء آخر يتغير، فأي حزمة برامج حاسوبية إحصائية سوف تتغلب بسهولة على التعقيدات الإضافية المرتبطة باستخدام توزيعات تي، ولهذا السبب سأناقش توزيع تي بشكل أكثر تفصيلًا في ملحق الفصل.

ولأننا حتى الآن باقون مع العينات الكبيرة (والتوزيع الطبيعي)، فإن أكثر الأمور جوهرية هو فهم ما للخطأ المعياري من أهمية. فكما هو الحال مع إجراء استطلاعات الرأي وغيرها من طرائق الاستدلال، لنا أن نتوقع وقوع أكثر من نصف معاملات الانحدار الملاحظة في نطاق خطأ معياري واحد من العامل المتغير للمجتمع الكامل*. وما نسبته نحو ٩٥ بالمائة سوف يقع في نطاق خطأين معياريين، وهلم جرًّا. وبهذا نكون شارفنا على الوصول إلى هدفنا، لأننا الآن قادرون على أن نقوم ببعض اختبارات الفرضيات. (أحقًّا ظننت أنك انتهيت من اختبار الفرضيات؟) فبمجرد أن يتاح لنا المعامل والخطأ المعياري، حينها نستطيع أن نختبر الفرضية المنعدمة التي تنص على عدم وجود علاقة بين المتغير المفسر والمتغير الاعتمادي (بمعنى أن الارتباط الحقيقي بين المتغيرين قدره صفر).

في ذلك المثال البسيط الذي أوردناه، مثال الطول والوزن، نستطيع أن نختبر أرجحية أن نجد كل بوصة في الطول ترتبط بـ ٤,٥ أرطال في عينة دراسة الحيوات المتغيرة، إذا لم يكن هناك بالفعل أي ارتباط بين الطول والوزن في المجتمع الأعم. لقد قمت بعمل تحليل انحدار مستخدمًا برنامجًا إحصائيًّا أوليًّا، فوجدت الخطأ المعياري لمعامل الطول قدره ١٣. وهذا يعني أننا إذا كررنا هذا التحليل – مع ١٠٠ عينة مختلفة مثلًا – فلنا أن نتوقع أن يكون معامل الانحدار الملاحظ واقعًا في نطاق خطأين معياريين من العامل المتغير الحقيقي للمجتمع كاملًا بنحو ٩٥ مرة من أصل ١٠٠.

ومن ثم فإننا قادرون على أن نعرض نتائجنا بطريقتين مختلفتين، وإن كانتا متصلتين إحداهما بالأخرى. أولًا نستطيع بناء فاصل ثقة قدره ٩٥ بالمائة. هنا نستطيع أن نقول إنه في ٩٥ مرة من أصل ١٠٠، لنا أن نتوقع فاصلنا للثقة، والذي هو ٤,٥ ± ٢٦، سيحتوي على العامل المتغير الحقيقي للمجتمع، والذي يتراوح ما بين ٤,٢٤ و٤,٧٦. وأي حزمة برامج إحصائية سوف تحسب لك فاصل الثقة كذلك. ثانيًا نستطيع أن نرى أن فاصل الثقة خاصتنا، الذي قدره ٩٥ بالمائة للمجتمع الحقيقي للارتباط الحقيقي بين الطول والوزن، لا يتضمن الصفر. ومن ثم فإننا قادرون على أن نرفض الفرضية المنعدمة التي تنص على عدم وجود ارتباط بين الطول والوزن في المجتمع الأعم عند مستوى ثقة قدره ٩٥ بالمائة. يمكن لهذه النتيجة أن تعرض على أنها ذات دلالة إحصائية عند مستوى٠٥. هناك احتمال قدره ٥ بالمائة أننا أخطأنا في رفض الفرضية المنعدمة.

الواقع أن نتائجنا أكثر تطرفًا من ذلك. فالخطأ المعياري (١٣) منخفض بشدة مقارنة بحجم المعامل (٤,٥). من الحقائق التي أتت بحكم التجربة أن المعامل يرجح

* «العامل المتغير في التجربة» هو تسمية منمقة لأي إحصائية تصف خصائص مجتمعٍ ما، فمتوسط وزن جميع البالغين من الرجال هو العامل المتغير لذلك المجتمع، وكذلك هو الانحراف المعياري. في هذا المثال يكون الارتباط الحقيقي بين الطول والوزن للمجتمع عاملًا متغيرًا لذلك المجتمع.

أن يصير ذا دلالة إحصائية إذا كان حجمه ضِعف حجم الخطأ المعياري*. كما أن حزمة البرامج الإحصائية سوف تحسب قيمة الــ p، والتي هي في هذه الحالة، مما يعني أن هناك احتمالًا قدره صفر لأن نخرج بناتج بذلك التطرف الذي لاحظناه (أو أكثر تطرفًا منه)، إذا لم يكن هناك ارتباط حقيقي بين الطول والوزن في المجتمع عامة. تذكر أننا لم نثبت أن طِوال القامة يزِنون أكثر في المجتمع الأعم، ولم نفعل سوى أن بيَّنا أن نتائجنا من عينة دراسة الحيوات المتغيرة ستكون شاذة جدًّا لو لم يكن هذا هو الحال.

تحليل الانحدار الأساسي الذي أجريناه نجده يُنتج إحصائية أخرى جديرة بالاعتبار، ألا وهي معامل تحديد الانحدار R^2، والذي يقيس إجمالي كمية التغايرات التي تفسرها معادلة الانحدار. إننا نعلم أن هناك نطاقًا واسعًا من التغايرات في الوزن في عينة دراسة الحيوات المتغيرة. فالكثير من الأشخاص الذين هم ضمن تلك العينة يزِنون أكثر من متوسط إجمالي المجموعة، والكثير منهم يزِنون أقل منه كذلك. يخبرنا الــ R^2 بكمٍّ التغاير الذي يقع حول المتوسط، والذي يرتبط بالفروق في الطول وحده. الإجابة في حالتنا هذه هي ٠.٢٥ أو ٢٥ بالمائة. لكن النقطة الأكثر أهمية هنا هي أن ٧٥ بالمائة من التغايرات في الوزن في عينتنا تظل غير ذات تفسير. فمن الواضح أن هناك عوامل غير الطول قد تساعدنا على إدراك أوزان المشاركين في دراسة الحيوات المتغيرة. هنا تغدو الأمور أكثر تشويقًا.

أعترف أنني استهللت هذا الفصل بأن سوَّقت لتحليل الانحدار كإكسير معجزة للبحث في العلوم الاجتماعية. لكن ما أنجزتُه حتى الآن هو أنني استخدمت حزمة البرامج الإحصائية ومجموعة مبهرة من البيانات لأُظهر أن طِوال القامة يزِنون أكثر من القصار. لعلك إن قمت بزيارة سريعة إلى مجمع تجاري لخلصت إلى نفس الحقيقة. والآن بعد أن فهمت الأساسيات، يمكننا أن نطلق عنان القوة الحقيقية لتحليل الانحدار. حان الوقت لننزع عن دراجتك عجلات تمرين المبتدئين.

كما وعدتك من قبل، فإن تحليل الانحدار سوف يسمح لنا بأن نكشف عن العلاقات المعقدة، والتي تؤثر فيها بعض العوامل على مخرج ما نحن معنيون به، كالدخل أو نتائج الاختبارات أو مرض القلب. عندما نُضَمِّن معادلة الانحدار متغيرات متعددة، حينها يمنحنا التحليل تقديرًا للارتباط الخطي بين كل متغير مفسر ومتغير اعتمادي، بينما يمسك بمتغيرات اعتمادية أخرى كثابت، أو «متحكم في» كل تلك العوامل الأخرى. لنبقَ مع الوزن لبرهة. لقد وجدنا ارتباطًا بين الطول والوزن، وإننا نعلم أن هناك عوامل أخرى يمكن أن تساعدنا على تفسير الوزن (العمر والجنس وممارسة الرياضة وهلم جرًّا). تحليل الانحدار (عادة ما يطلق عليه تحليل الانحدار المتعدد أو تحليل الانحدار متعدد المتغيرات،

---

* عندما تنص الفرضية المنعدمة على أن معامل الانحدار يساوي صفرًا (كما هو معتاد في هذه الحالة)، حينها تعرف نسبة معامل الانحدار إلى الخطأ المعياري باسم إحصائية تي. وذلك سوف يوضح أيضًا في ملحق الفصل.

إذا ما انطوى على أكثر من متغير مفسر) سوف يمنحنا معاملًا لكل متغير مفسر مضمَّن في معادلة الانحدار. بعبارة أخرى ما العلاقة بين الطول والوزن بين أولئك الذين هم بنفس الطول ومن نفس الجنس؟ ما إن نمتلك أكثر من متغير مفسر، فلا يتسنى لنا أن ننشر تلك البيانات في اثنين من الأبعاد. (حاول أن تتخيل رسمًا بيانيًّا يمثل الوزن والجنس والطول والعمر لدى كل مشارك في دراسة الحيوات المتغيرة). لكن المنهجية الأساسية هي نفسها كما في مثال الطول والوزن البسيط. وكلما أضفنا متغيرات مفسرة، قامت حزمة البرامج الإحصائية بحساب معامل الانحدار لتقليص إجمالي المتبقيات المربعة لمعادلة الانحدار.

لنعمل ببيانات دراسة تغير الحيوات الآن، ثم بعدها سوف أعود بك وأقدِّم لك تفسيرًا منطقيًّا لكيفية عمل ذلك الانقسام الإحصائي الذي هو أشبه بانفلاق البحر الأحمر إلى فلقين. يمكننا أن نبدأ بإضافة متغير واحد إضافي إلى المعادلة التي تفسر أوزان المشاركين في تلك الدراسة، ألا وهو العمر. عندما نجري تحليل انحدار يتضمن الطول والعمر كمتغيرين مفسرين للوزن، فذلك الذي ستحصل عليه:

الوزن = - ١٤٥ + ٤٫٦ × (الطول بالبوصات) + ١ × (العمر بالسنين)

المعامل للعمر هو ١ يمكن أن يفسر هذا على أن كل سنة إضافية في العمر ترتبط بـ ١ رطل إضافي في الوزن، مع الاحتفاظ بالطول كثابت. في أي مجموعة من الناس المتساوين طولًا سَيَزِن الأشخاص الأكبر عمرًا بعشر سنين رطلًا أكثر من غيرهم تقريبًا. ليس ذلك بالتأثير الكبير لكنه يتسق مع ما نلاحظه في حياتنا. يصير المعامل ذو دلالة عند مستوى ٠٫٠٥.

لعلك قد لاحظت أن معامل الطول قد ازداد زيادة طفيفة. ما إن نُدخل العمر ضمن تحليل الانحدار، حتى يتكون لدينا فهْم أكثر صَقْلًا للعلاقة بين الطول والوزن. من بين أولئك الأشخاص في عيِّنتنا الذين هم في نفس العمر أو «يحتفظون بالعمر كثابت»، ترتبط كل بوصة إضافية في الطول بـ ٤٫٦ أرطال في الوزن.

لنُضِفْ متغيرًا واحدًا آخر، ألا وهو الجنس. هنا سيختلف الحال قليلًا لأن للجنس احتمالين فقط، فإما ذكر أو أنثى. فكيف للمرء أن يضع في تحليل الانحدار كلمة ذكر أو كلمة أنثى؟ الحل هو أن نستخدم ما يسمى بالمتغير الثنائي أو المتغير الوهمي. ففي مجموعة البيانات التي لدينا نقوم بإدخال رقم ١ للمشاركات الإناث و٠ للمشاركين الذكور. (لا يقصد بذلك أن يكون حكمًا تقديريًّا). ومن ثم يمكن تفسير معامل الجنس على أنه تأثير كون المشاركة أنثى على الوزن مع ثبات بقية العوامل. المعامل هنا هو – ٤٫٨، وهو الأمر غير المفاجئ. يمكن أن نفسر ذلك على أنه بين الأشخاص الذين هم بنفس الطول والعمر، تزن النساء ٤٫٨ أرطال أقل من الرجال. والآن بدأنا قدرًا من قوة تحليل الانحدار المتعدد. نعرف أن النساء بشكل عام أقصر قامة من الرجال، لكن معاملنا قد وضع ذلك

في الاعتبار بما أننا ثبَّتنا عامل الطول. أما ما عزلنا هنا هو تأثير كون المشارِكة أنثى. هنا تصير معادلة الانحدار الجديدة:

الوزن = - ١١٨ + ٤,٣ × (الطول بالبوصات) + ١٢. (العمر بالسنين) - ٤,٨ (إذا كان الجنس أنثى)

أفضل تقديراتنا لوزن امرأة في الثالثة والخمسين من عمرها يبلغ طولها ٥ أقدام و٥ بوصات هو: - ١١٨ + ٤,٣ (٦٥) + ١٢. (٥٣) - ٤,٨ = ١٦٣ رطلًا.

وأفضل تخميناتنا لوزن رجل في الخامسة والثلاثين من عمره يبلغ طوله ٦ أقدام و٣ بوصات - ١١٨ + ٤,٣ (٧٥) + ١٢. (٣٥) = ٢٠٩ أرطال. لقد تخطينا آخر وصف في نتائج تحليل الانحدار ذلك (- ٤,٨) لأن ذلك المشارك ذكر لا أنثى.

الآن نستطيع أن نختبر أمورًا أكثر تشويقًا وأقل قابلية لأنْ يُتنبأ بها. ماذا عن مستوى التعليم؟ كيف لهذا أن يؤثر على الوزن؟ سوف أفترض أن الأفراد الأكثر تعليمًا يتمتعون بوعي صحي أكبر، ومن ثم فأوزانهم أقل مع تثبيت العوامل الأخرى. كما أننا لم نختبر أي مقياس لممارسة الرياضة، لنفترض مع تثبيت العوامل الأخرى، أن من يمارسون الرياضة سيكونون أقل وزنًا.

ماذا عن الفقر؟ هل لكون المرء من ذوي الدخل المنخفض في أمريكا تأثيرات تفصح عن نفسها في وزنه؟ لقد تقصَّت دراسة الحيوات المتغيرة عما إذا كان المشاركون يتلقون قسائم طعام مجاني، وهو الأمر الذي يعد مقياسًا جيدًا للفقر في أمريكا. وأخيرًا أنا مهتم بالعرق. نعلم أن ذوي البشرة الملونة يعيشون خبرات حياتية مختلفة في الولايات المتحدة بسبب عرقياتهم. هناك عوامل ثقافية وإسكانية ترتبط بالعرق في أمريكا، وهي العوامل ذات المضامين المتعلقة بالوزن. ما زالت هنالك الكثير من المدن التي تتميز بدرجات عالية من الفصل العنصري، والأمريكان ذوو الأصول الأفريقية هم الأكثر عرضة من بين جميع المواطنين لأن يعيشوا في «صحارى غذائية»، والتي هي مناطق تقل فيها محالُّ البقالة التي تبيع الفواكه والخضروات وغيرها من الأغذية الطازجة.

يمكننا أن نستخدم تحليل الانحدار لفصل كل تأثير مستقل لكل عامل من العوامل التي يحتمل أن تكون مفسرة والتي وصفناها فيما سبق. فمثلًا نستطيع أن نعزل الارتباط بين العرق والوزن، مع تثبيت عوامل اجتماعية واقتصادية أخرى كالخلفية التعليمية والفقر. فما الارتباط الإحصائي بين الوزن وكون المرء أسود البشرة بين مَن تخرجوا من المدرسة الثانوية وهم في نفس الوقت يستحقون تلقِّي قسائم الطعام المجاني؟

عند تلك النقطة تصير معادلة الانحدار التي لدينا طويلة جدًّا، للدرجة التي ستصير معها طباعة كامل النتائج أمرًا مرهقًا. وعادة ما تقوم البحوث الأكاديمية بإدراج جداول

كبيرة تلخص نتائج معادلات الانحدار المتعددة. ولقد ضمَّنتُ ملحق الفصل جدولًا يعرض للنتائج الكاملة لمعادلة الانحدار. لكن لنعرض الآن أهم ما سيحدث عندما نضيف إلى معادلتنا التعليم وممارسة الرياضة والفقر (قياسًا باستحقاق قسائم الطعام المجاني) والعرق.

لا تزال متغيراتنا الأصلية كلها (الطول والعمر والجنس) تحمل دلالة. لكن المعاملات تتغير طفيفًا بإضافتنا للمتغيرات المفسرة. تصير متغيراتنا الجديدة كلها ذات دلالة إحصائية عند مستوى ٠.٠٥. وقد قفز الـ R^2 للانحدار من ٢٥. إلى ٢٩. (تذكر أن أي R^2 قدره صفر يعني أن معادلتنا للانحدار لا تفعل خيرًا مما يفعله المتوسط في التنبؤ بوزن أي فرد من أفراد العينة، وأن أي R^2 قدره ١ يعني أن معادلة الانحدار تتنبأ بشكل صحيح تمامًا بوزن كل فرد من أفراد العينة). لكن كثيرًا من التغاير الحادث في أوزان أفراد العينة يظل غير مفسر.

كما افترضت أنا، اتضح أن التعليم ذو ارتباط سلبي بالوزن. فبين المشاركين في دراسة الحيوات المتغيرة ترتبط كل سنة قضاها المرء في التعليم بسبعة أرطال سالبة – ٧.

ولم يكن مثيرًا للدهشة أن ترتبط ممارسة الرياضة بالوزن ارتباطًا سالبًا. تضمنت دراسة الحيوات المتغيرة فهرسًا يقيم كل مشارك في الدراسة طبقًا لمستواه أو مستواها في النشاط البدني. والأشخاص الذين وقعوا في الخُمس الأخير من مسلسل من يمارسون النشاط البدني يزِنون نحو ٣,٧ أرطال أكثر من بقية البالغين في العينة مع تثبيت بقية العوامل. ونفس أولئك الذين في الخُمس الأخير من مسلسل من يمارسون النشاط البدني يزِنون أكثر من أولئك البالغين الذين هم في الخُمس الأول بنحو ٩ أرطال.

والأفراد الذين يتلقون قسائم الطعام المجاني (المؤشر على الفقر في تحليل الانحدار هذا) هم أثقل وزنًا من بقية البالغين. يزن متلقُّو قسائم الطعام المجاني من المشاركين في دراسة الحيوات المتغيرة نحو ٥,٦ أرطال أكثر من بقية البالغين في الدراسة مع تثبيت بقية العوامل.

لقد اتضح أن متغير العرق متغير مثير للاهتمام. فحتى عند التحكم في جميع المتغيرات الأخرى التي وصفناها حتى هذه اللحظة، يظل العرق ذا أهمية كبيرة عند تفسير الوزن. في دراسة الحيوات المتغيرة يزن البالغون سُودُ البشرة من غير الهسبانيين نحو ١٠ أرطال أكثر مما يزن بقية البالغين في الدراسة. وعشرة أرطال هو وزن كبير سواء بشكل عام أو مقارنة بتأثيرات المتغيرات المفسرة الأخرى في معادلة الانحدار. وليس ذلك بشذوذ في البيانات. تقدر قيمة الـ P في المتغير الوهمي للسود من غير الهسبانيين بـ ٠٠٠. ويمتد مجال الثقة المقدر بـ ٩٥ بالمائة من ٧,٠ أرطال إلى ١٢,٣ رطلًا.

ماذا يحدث؟ بكل أمانة ليست لديَّ أيُّ فكرة. دعني أُعِدْ عليك فكرة جاءت مطمورة في وقت سابق في إحدى الحواشي السفلية: إنني هنا لا أفعل سوى أن أعبث بالبيانات كي أُظهر لك كيفية عمل تحليل الانحدار. والتحليلات الواردة هنا بالنسبة إلى الأبحاث الأكاديمية الحقيقية هي كرياضة الهوكي التي تُلعب في الشوارع نسبة إلى البطولة الوطنية للهوكي. فلو كان هذا مشروعًا بحثيًّا حقيقيًّا لتطلب أسابيع أو أشهرًا من التحليل اللاحق لاستكشاف هذه النتيجة. ما أستطيع قوله هو أنني قد وضحت السبب في كون تحليلات الانحدار المتعددة هي الأداة الفضلى التي في حوزتنا للعثور على الأنماط ذات المعنى في مجموعات البيانات الكبيرة المعقدة. بدأنا بمثال شديد الابتذال يقيِّم كميًّا العلاقة بين الطول والوزن، ولم يمضِ الكثير من الوقت حتى وجدنا أنفسنا نخوض حتى رُكَبِنا في قضايا ذات دلالات اجتماعية حقيقية.

في نفس الإطار يمكنني أن أعرض عليك دراسة حقيقية استخدَمت تحليل الانحدار لاستكشاف قضية ذات دلالة اجتماعية، ألا وهي التمييز في محال العمل على أساس الجنس. ما يثير الفضول في مسألة التمييز هو أنه من الصعب ملاحظته بشكل مباشر. فليس من رب عمل يصرح علانية بأن أحدًا من موظفيه يتقاضى أقل من زملائه بسبب عرقه أو جنسه، أو أنه رفض توظيف أحدهم لأسباب تمييزية (وهو الأمر الذي سيؤدي بهذا الشخص لأن يظل في وظيفة أخرى ذات راتب متدنٍّ على الأرجح). لكن ما نلاحظه هو وجود فجوات في الرواتب قائمة على العرق أو الجنس، والتي قد تكون نتاجًا لتمييز ما: كأن يتقاضى البيض أكثر مما يتقاضاه السود، أو يتقاضى الرجال أكثر من النساء، وهلم جرًّا. لكن التحدي المنهجي يكمن هنا في كون تلك الفجوات الملاحظة قد تكون نتاجًا لفروق تتعلق بالعاملين أنفسهم، وليست من قبيل التمييز في محل العمل، كحقيقة أن النساء يَمِلْن لاختيار الدوام الجزئي في العمل. كم من هذه الفجوة في الأجور ترتبط بالإنتاجية في العمل، وكم منها يعزى لتمييز في القوة العاملة؟ لا يمكن أن يدَّعي أحدهم أن هذا سؤال تافه.

يستطيع تحليل الانحدار أن يساعدنا على الإجابة عنه، لكن منهجيتنا هذه المرة ستكون أكثر التواءً من التحليل الذي أجريناه لتفسير الوزن. فبما أننا لا نستطيع أن نقيس التمييز بشكل مباشر، فسوف نقوم باختبار عوامل أخرى تفسر الأجور بشكل تقليدي، كالتعليم والخبرة والمجال الوظيفي، وهلم جرًّا. فالتمييز هو قضية ظرفية، بمعنى أننا لو وجدنا تلك الفجوة الكبيرة في الأجور لا تزال قائمة، حتى بعد تثبيت عوامل أخرى عادة ما تفسر الأجور، فإن التمييز يكون هنا متهمًا. وكلما زادت حصة الأجور غير المفسرة، ازداد الأمر إثارة للريبة. كمثال على ذلك لننظر في تلك الدراسة التي قام بها ثلاثة من الباحثين في الاقتصاد، والذين درسوا سلم الأجور لعينة قوامها ٢٥٠٠ رجل وامرأة حاصلين على ماجستير في إدارة الأعمال من مدرسة بوث للأعمال في جامعة شيكاجو[١] وقت تخرج

هؤلاء الأشخاص كان متوسطَي الرواتب الأولى التي تقاضاها الرجال والنساء متقاربَين؛ إذ بلغ ١٣٠٠٠٠ دولار للرجال، و١١٥٠٠٠ للنساء. لكن بعد قضاء عشرة أعوام على ذمة القوة العاملة حدثت فجوة كبيرة بين الرقمين، فصارت النساء يتقاضين رواتب أزهد من زملائهن الرجال بنسبة قدرها ٤٥ بالمائة، ٢٤٣٠٠٠ دولار أمريكي في مقابل ٤٤٢٠٠٠ دولار أمريكي. وفي عينة أكبر قوامها أكثر من ١٨٠٠٠ حاصل وحاصلة على ماجستير إدارة الأعمال من الذين انضموا إلى القوة العاملة ما بين عامي ١٩٩٠ و٢٠٠٦، وُجد أن الإناث يتقاضين رواتب أقل بنسبة ٢٩ بالمائة. ماذا يحل بالنساء بعد أن ينضممن إلى القوة العاملة؟

يقول كاتبو الدراسة (ماريان برتراند من مدرسة بوث للأعمال وكلاوديا جولدنج ولورنس كاتز من هارفارد)، إن التمييز ليس بالتفسير الراجح للقدر الأكبر من هذه الفجوة. فعند إضافة الباحثين للمزيد من المتغيرات المفسرة للتحليل، تتلاشى تلك الفجوة في الأجور، والقائمة على الجنس. فمثلًا يلتحق الرجال بصفوف متخصصة في التمويل أكثر من النساء ضمن برنامج ماجستير إدارة الأعمال، ويتخرجون بتقديرات دراسية أعلى في المتوسط. عندما تدرج تلك البيانات كمتغيرات تحكم في معادلة الانحدار، تتقلص الحصة غير المفسرة من الفجوة بين ما يتقاضاه الرجال والنساء بنسبة ١٩ بالمائة. وعندما تضاف متغيرات مفسرة متعلقة بخصائص أخرى للعمل، كنوع رب العمل وعدد ساعات العمل، تتقلص الحصة غير المفسرة للفجوة القائمة على جنس الموظف إلى ما دون ٤ بالمائة.

بالنسبة إلى العاملين الذي قضوا أكثر من عشر سنين ضمن قوة العمل، يستطيع الباحثون أن يأتوا بتفسير قاطع لتلك الفجوة الجنسية في الأجور من خلال العوامل غير ذات الصلة بالتمييز في العمل إلا فيما نسبته ١ بالمائة من تلك الفجوة*. وقد خلصوا من الدراسة إلى «أننا نستطيع تحديد ثلاثة أسباب تقريبية لهذه الفجوة الواسعة والآخذة في الاتساع في الأجور والقائمة على الجنس، وهي: الفروق في مستوى التدريب الذي يتلقاه الموظفون قبل الحصول على ماجستير إدارة الأعمال، والفروق من حيث حدوث توقفات مؤقتة في المسيرة المهنية، والفروق في عدد ساعات العمل. تلك المحددات الثلاثة قادرة على تفسير مجمل الفروق الجنسية خلال السنين التي تلي نَيْل ماجستير إدارة الأعمال».

آمل أن أكون قد أقنعتك بقيمة تحليلات الانحدار المتعددة، وعلى وجه التحديد تلك الرؤى البحثية التي انبثقت من قدرتنا على عزل تأثير متغير مفسر واحد، في الوقت الذي نثبت فيه العوامل الأخرى المشوشة. إنني لم أقدم لك حتى الآن تفسيرًا منطقيًّا لكيفية

* قد تؤثر القوى التمييزية في المجتمع على المهن التي تختارها النساء أو على حقيقة أنهن أكثر قابلية من الرجال لأن يقطعن مسيرتهن الوظيفية كي يَرْعَيْنَ أطفالهن. لكن هذه القضايا المهمة هي قضايا مستقلة عن ذلك السؤال المحدود عما إذا كانت النساء يتقاضين أكثر من زملائهن الرجال في ذات الوظيفة.

عمل ذلك؛ «الإكسير الإحصائي المعجزة». عندما نستخدم تحليل الانحدار لتقييم العلاقة بين التعليم والوزن مع تثبيت العوامل الأخرى، فكيف تتحكم حزمة البرامج الإحصائية في عوامل مثل الطول والجنس والعمر والدخل عندما نعلم أن المشاركين في دراسة الحيوات المتغيرة ليسوا سواءً في تلك الاعتبارات؟

كي أقرب إلى ذهنك كيف يمكن عزل تأثير متغير وحيد في الوزن، وليكن التعليم مثلًا، فلتتخيل الموقف التالي. لنفترض أن جميع المشاركين في دراسة الحيوات المتغيرة مجموعون في مكان واحد، وليكن ذلك المكان فارمنجهام بماساتشوستس. والآن لنفترض أنه قد تم فصل الرجال عن النساء. ثم لنفترض بعدها أنه قد تم تقسيم كلٍّ من الرجال والنساء طبقًا لأطوالهم. فتصير هناك غرفة بها رجال يبلغون الست أقدام طولًا. وإلى جوارها غرفة أخرى بها رجال طولهم ست أقدام وبوصة واحدة، وهلم جرًّا بالنسبة إلى كلا الجنسين. ولو توافر لدينا العدد الكافي من المشاركين في الدراسة، نستطيع حينها أن نقوم بتقسيمات فرعية لكل غرفة طبقًا لدخل الفرد. في النهاية سنجد لدينا الكثير من الغرف التي تحتوي كلٌّ واحدة منها على أفراد متطابقين في كل شيء عدا التعليم والوزن، وهما المتغيران اللذان نُعنى بهما. ستكون هناك غرفة بها رجال في الخامسة والأربعين من عمرهم ويبلغ طولهم خمس أقدام وخمس بوصات ويتقاضون ما بين الـ ٣٠٠٠٠ والـ ٤٠٠٠٠ دولار أمريكي سنويًّا. وفي الغرفة المجاورة لها ستكون نساء في الخامسة والأربعين من عمرهن يبلغن خمس أقدام وخمس بوصات طولًا ويتقاضين ما بين الثلاثين ألفًا والأربعين ألف دولار أمريكي سنويًّا. وهلم جرًّا (وجرًّا وجرًّا).

لكن سيظل هناك بعض التغاير في الوزن داخل كل غرفة، إذ سيظل مَن هم مِن نفس الجنس وبنفس الطول ويحصلون على نفس الدخل تتباين أوزانهم – وإن كان التباين داخل كل غرفة سيفترض أنه أقل كثيرًا من التباين على مستوى العينة كلها. هدفنا الآن أن نرى كم من ذلك التغاير الباقي مرده إلى التعليم. بعبارة أخرى، ما العلاقة الخطية الفضلى بين التعليم والوزن في كل غرفة؟

لكن التحدي الأخير يكمن في كوننا لا نريد معاملات مختلفة في كل «غرفة». المقصد من هذا التمرين هو حساب معامل مفرد هو الأفضل في التعبير عن العلاقة بين التعليم والوزن في العينة كلها مع تثبيت العوامل الأخرى. ما نريد حسابه هنا هو معامل التعليم المفرد الذي نستطيع استخدامه في كل غرفة لتقليص مجموع المتبقيات المربعة للغرف كلها مجتمعة. أي إن معامل التعليم يقلص مربع الوزن غير المفسر لكل فرد في الغرف كلها. يصير هذا معامل الانحدار لأنه التفسير الأفضل للعلاقة الخطية بين التعليم والوزن لهذه العينة عندما نبقي الجنس والطول والدخل كثوابت.

ملاحظة جانبية هنا هو أنك الآن تستطيع أن ترى لماذا هي مفيدة جدًّا مجموعات

البيانات الكبيرة تلك. فهي تسمح لنا بأن نحكم الكثير من العوامل في نفس الوقت الذي نجد فيه الكثير من المشاهدات في كل «غرفة». ومن الواضح أن أي حاسوب قادر على أن يفعل كل ذلك في جزء من الثانية دون الحاجة إلى الدفع بآلاف البشر إلى غرف مختلفة.

لنُنهِ الفصل كما بدأناه بالصلة بين التوتر في العمل ومرض شرايين القلب التاجية. كانت دراسات وايت هول التي أُجريت على الموظفين المدنيين البريطانيين قد نشدت قياس الارتباط بين الدرجة الوظيفية والوفاة من جراء مرض الشرايين القلبية في السنين التالية لتولي الوظيفة. واحدة من الدراسات المبكرة تتبعت ١٧٥٣٠ من العاملين المدنيين على مدى سبعة أعوام ونصف[٢]. وخلص الباحثون إلى أن «الرجال العاملين بالوظائف الدنيا كانوا أقصر قامة وأثقل وزنًا مما يتناسب مع طولهم، وذوي ضغط دم أعلى، ومعدلات أعلى لجلوكوز البلازما، ويدخنون أكثر، ويقضون وقتًا أقل في ممارسة النشاط البدني في أوقات الفراغ؛ من الرجال الذين يتولون الوظائف العليا. لكن رغم هذا فعندما تم تحديد نصيب كل تلك العوامل، مضافًا إليها كولسترول البلازما، من التأثير على معدل الوفيات، وجد أن الارتباط العكسي الحادث بين الدرجة الوظيفية و[مرض الشرايين القلبية] ما زال قويًّا». ذلك «النصيب» الذي يشيرون إليه والخاص بعوامل مخاطرة أخرى قد حددته وسائل لتحليل الانحدار*. تُظهر الدراسة أنه بتثبيت العوامل الصحية الأخرى (ومن بينها الطول الذي هو مؤشر مناسب على الصحة والتغذية اللذين تمتع الفرد بهما خلال طفولته)، فإن العمل في وظائف «دنيا» يمكنه أن يقتل المرء.

دائمًا ما يكون التشكك استجابة أولى جيدة. كتبتُ في بداية الفصل أن الوظائف ذات «التحكم المحدود» ضارة بصحتك. وذلك قد يترادف أو لا يترادف مع أن يكون المرء في درجة منخفضة من سلم المديرين. دراسة لاحقة استخدمت عينة ثانية قوامها ١٠٣٠٨ من العاملين المدنيين البريطانيين حاولت أن تنقِّب عن الفرق بين الاثنين[٣]. مجددًا تم تقسيم العاملين طبقًا للدرجات الإدارية – عالية ومتوسطة ودنيا – لكن هذه المرة تم إعطاء المشاركين استقصاءً مكونًا من خمسين سؤالًا يقيِّم «مدى تحكمهم في اتخاذ القرارات». وقد تضمن الاستقصاء أسئلة على غرار «هل تملك اختيار الطريقة التي تؤدي بها عملك؟» واختيارات

* تختلف هذه الدراسات اختلافًا طفيفًا عن معادلات الانحدار التي عرضناها في بداية هذا الفصل. وفي هذه الدراسات يكون المخرج المنشود أو المتغير الاعتمادي ثنائيًّا. فالمشارك في الدراسة إما يعاني نوعًا من أنواع المشكلات الصحية المتعلقة بالقلب خلال فترة الدراسة وإما لا يعانيها. ونتيجة لهذا يستخدم الباحثون أداة تسمى الانحدار اللوجستي المتعدد المتغيرات. الفكرة الأساسية هنا هي ذاتها فكرة نماذج المربعات الصغرى العادية التي بيَّنَّاها في هذا الفصل. فكل معامل يعبر عن تأثير متغير مفسر بعينه على المتغير الاعتمادي مع تثبيت تأثيرات المتغيرات الأخرى في ثابت النموذج. لكن الاختلاف الأساسي هنا هو أن متغيرات المعادلة كلها تؤثر على أرجحية وقوع حدثٍ ما، كالإصابة بأزمة قلبية خلال فترة الدراسة. فمثلًا في هذه الدراسة يكون العاملون في الوظائف الدنيا أكثر عرضة للإصابة بـ «حدث متعلق بالشرايين التاجية» خلال فترة الدراسة ١٫٩٩ مرة أكثر من العاملين بالوظائف العليا بعد التحكم في جميع عوامل الخطورة المتعلقة بمخاطر الإصابة بأزمة في الشرايين التاجية.

فئوية في الإجابة (تتدرج من «أبدًا» إلى «عادة») إلى إفادات على غرار «أستطيع أن أقرر متى أحصل على استراحة». وجد الباحثون أن العاملين «الأقل تحكمًا» أكثر عرضة بشكل كبير لأن يصابوا بمرض الشرايين التاجية على المدى الزمني للدراسة من العاملين «الأكثر تحكمًا». لكن الباحثين وجدوا كذلك أن العاملين بوظائف ذات متطلبات قاسية ليسوا في محل أكثر خطورة لأن يصابوا بمرض القلب، ومثلهم العاملون الذين لا يتمتعون إلا بالقليل من الدعم المجتمعي في وظائفهم. فالافتقار إلى التحكم يبدو قاتلًا بالمعنى الحرفي للكلمة.

لدراسات وايت هول خاصيتان ترتبطان عادة بالأبحاث القوية. أولاهما أن نتائجها قد استنسخت في أماكن أخرى. فقد تطورت فكرة «التحكم الضعيف» في أدبيات الصحة العامة لتصير مصطلحًا هو «الإجهاد الوظيفي»، والذي يَسِمُ تلك الوظائف ذات «المتطلبات المهنية النفسية القاسية» و«النطاق المحدود لاتخاذ القرارات». ما بين عامي ١٩٨١ و١٩٩٣ نُشرت ست وثلاثون دراسة حول هذا الموضوع، وخلصت معظمها إلى وجود ارتباط إيجابي بين الإجهاد الوظيفي ومرض القلب[٤].

وثانيتهما أن الباحثين قد بحثوا عن دلائل بيولوجية معززة لتفسير الآليات التي بها يسبب هذا النوع من التوتر اعتلال الصحة، وقد وجدوا تلك الدلائل. فظروف العمل التي تتضمن متطلبات قاسية مع مستوًى ضعيف من التحكم يمكن أن تؤدي إلى ردات فعل فسيولوجية (كانطلاق هرمونات ذات علاقة بالتوتر) تزيد من مخاطر الإصابة بمرض القلب على المدى الطويل. وحتى الأبحاث التي أجريت على الحيوانات قد لعبت دورًا في هذا الصدد، فالقرود الأدنى مرتبة وقرود البابون (التي تحمل شبهًا بالعاملين المدنيين القابعين في أسفل سلم السلطة) بها اختلافات فسيولوجية عن أقرانها ذا المراتب العليا، وهي اختلافات تجعلها أكثر عرضة لأمراض الشرايين القلبية[٥]

في كل ما سبق لا يهمنا إلا حقيقة أنه لا يجدر بك أن تصير قرد بابون متدني الرتبة، وهي الفكرة التي لا أفتأ أذكِّر بها أطفالي كلما عنَّ لي ذلك، وخاصة ابني. أما الرسالة الأكبر هنا فهي أن تحليل الانحدار هو كما يزعم البعض أكثر الأدوات أهمية، والتي يستخدمها الباحثون لتحديد الأنماط ذات المعنى في مجموعات البيانات الكبيرة. إننا في الظروف العادية لا نستطيع أن نجري تجارب ضابطة لنعرف عن التمييز الوظيفي أو العوامل التي تسبب مرض القلب. إن ما فَطِنَّا إليه عن هذه القضايا الاجتماعية المهمة، وغيرها كثير، قد أتت من هذه الأدوات الإحصائية التي تناولناها في هذا الفصل. بل إننا لن نكون مبالغين لو قلنا إن شطرًا كبيرًا من الأبحاث العلمية المهمة التي أجريت في العلوم الاجتماعية خلال نصف القرن الماضي (وبالأخص بعد ظهور الحواسيب القوية الزهيدة الثمن) قد قام على تحليل الانحدار.

تحليل الانحدار يضخم من حجم المنهج العلمي، فبسببه صرنا أكثر صحة وأمنًا وثقافة.

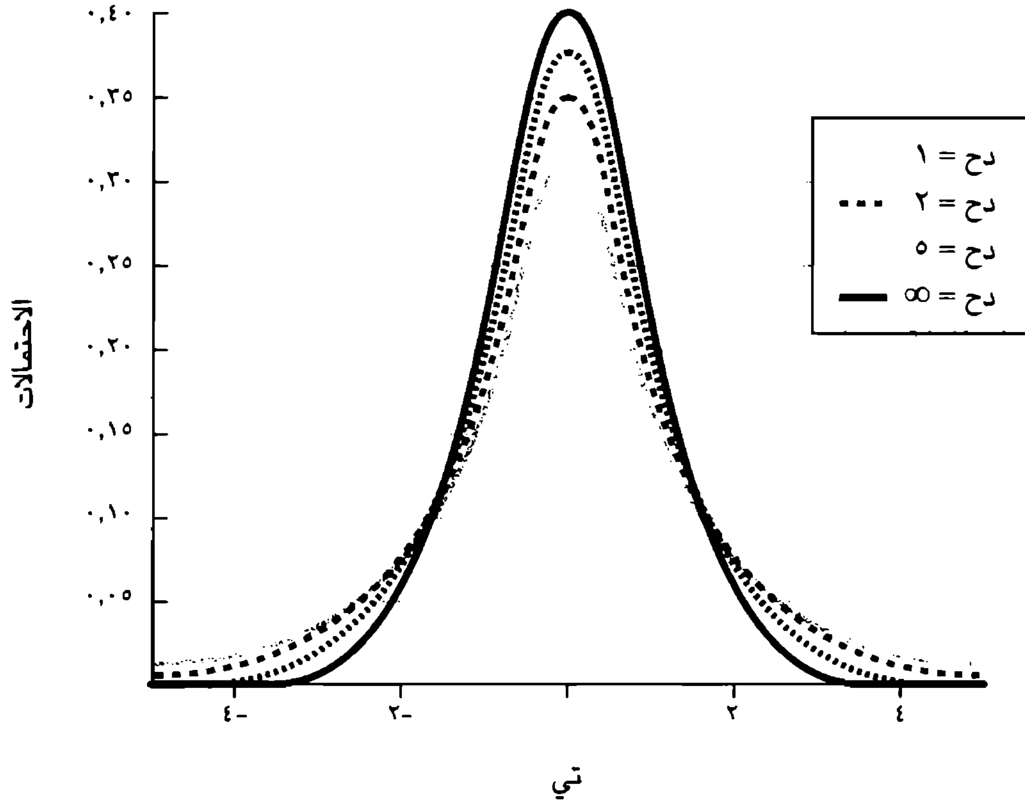
إذن ما الذي قد يَنْحُو منحًا خطأً عند التعامل مع تلك الأداة القوية المذهلة؟ فلتواصل القراءة.

ملحق الفصل الحادي عشر

توزيع تي.

تغدو الحياة أكثر مخاتلة عندما نجري تحليل الانحدار (أو غيره من طرائق الاستدلال الإحصائي) على عينة بيانات صغيرة. لنفترض أننا كنا نحلل العلاقة بين الطول والوزن استنادًا إلى عينة قوامها ٢٥ بالغًا فقط، لا مجموعة هائلة من البيانات كتلك التي حصلنا عليها من دراسة الحيوات المتغيرة. يقتضي المنطق أن نكون أقل ثقة في أن نعمم نتائجنا التي خرجنا بها من هؤلاء الأشخاص الخمسة وعشرين، لا من عينة قوامها ٣٠٠٠ شخص على كامل مجتمع البالغين. واحدة من الفِكَر الرئيسية التي امتدت على مدى هذا الكتاب هي أن العينات الأصغر تخلق المزيد من التشتت في المخرجات. لكن عينتنا ذات الخمسة والعشرين فردًا سوف تمنحنا، مع ذلك، معلومات ذات معنًى تمامًا كما ستفعل عينات مكونة من خمسة أشخاص أو عشرة – لكن ما قدر هذا المعنى؟

يجيب توزيع تي. عن هذا السؤال. فلو قمنا بتحليل الارتباط بين الطول والوزن في عينات مكررة قوام الواحدة منها خمسة وعشرون شخصًا، فلن يعود باستطاعتنا أن نفترض أن معاملات الطول المتعددة التي سنحصل عليها سوف تتوزع بشكل طبيعي حول المعامل «الحقيقي» للطول في كامل المجتمع. سوف تتوزع حول المعامل الحقيقي للمجتمع كله، لكن توزيعها لن يتخذ شكل المنحنى الجرسي الطبيعي الذي أَلِفناه. لكننا سوف نفترض أن تكرار العينات المكونة من خمسة وعشرين فردًا سوف يؤدي إلى المزيد من التشتت حول معامل المجتمع الحقيقي – ومن ثم سينتُج توزع ذو «أذيال أكثر سُمكًا». ولو كررنا عينات مؤلفة من ١٠ أشخاص لازداد التشتت أكثر ولنتجت أذيال أكثر سُمكًا. وتوزيع تي هو في الحقيقة سلسلة أو «عائلة» من وظائف قدر الاحتمالات والتي تتباين طبقًا لحجم العينة. أو على وجه التحديد كلما زادت البيانات في عينتنا زادت «درجات الحرية» لدينا في تحديد التوزيع الأنسب الذي منه نقيِّم نتائجنا. وإذا ما التحقت بصف إحصاء أكثر تقدمًا، لتعلمت كيف تحسب درجات الحرية، والتي هي تساوي تقريبًا عدد المشاهدات في عينتنا. فمثلًا سيكون لأي تحليل انحدار أساسي بعينة مكونة من ١٠ أشخاص ومتغير مفسر واحد، ٩ درجات من الحرية. وكلما زادت درجات الحرية لدينا، ازددنا ثقة في أن عينتنا تمثل المجتمع الحقيقي، و«ضاق» توزيعنا كما يظهر الرسم البياني التالي.



عندما يعلو رقم درجات الحرية، حينها يتحول توزيع تي إلى توزيع طبيعي. ولهذا فإننا عندما نعمل على مجموعات بيانات كبيرة نستطيع أن نستخدم التوزيع الطبيعي في حساباتنا المتنوعة.

لا يفعل توزيع تي سوى أنه يضيف فارقًا بسيطًا لنفس عملية الاستدلال الإحصائي الذي ظللنا نستخدمه على مدى كامل الكتاب. إننا ما زلنا نشكل فرضية منعدمة ثم نختبرها في مقابل بعض البيانات التي نلاحظها. وإذا ما كانت تلك البيانات التي لاحظناها مستبعدة جدًّا لو تحققت الفرضية المنعدمة، حينها سوف نرفض تلك الفرضية المنعدمة. الأمر الوحيد الذي يتغير مع توزيع تي هو الاحتمالات التي تترتب عليه لتقييم المخرجات المشاهدة. كلما «زاد سُمك» الذيل في توزيع احتمالات بعينة (مثل توزيع تي لثماني درجات من الحرية)، زاد التشتت الذي يجدر بنا أن نتوقعه من بياناتنا المشاهدة بمحض الصدفة، ومن ثم صرنا أقل ثقة في رفضنا للفرضية المنعدمة.

فمثلًا لنفترض أننا نجري معادلة انحدار، وأن الفرضية المنعدمة هي أن معامل أحد المتغيرات يبلغ صفرًا. ما إن نحصل على نتائج الانحدار حتى يتسنى لنا أن نحسب إحصائية تي، والتي هي نسبة المعامل المشاهد للخطأ المعياري لذلك المعامل*. ثم يتم

* المعادلة الأكثر عمومية لحساب إحصائية تي هي كما يلي:

$$t_b = \frac{b - b_o}{SE_b}$$

حيث b هو المعامل الملاحظ وb_o هو الفرضية المنعدمة لذلك المعامل وSE_b هو الخطأ المعياري للمعامل الملاحظ b.

تقييم إحصائية تي تلك في مقابل توزيع تي المناسب لذلك الحجم من عينة البيانات (بما أن هذا هو ما يلعب الدور الأكبر في تحديد درجات الحرية). عندما تكون إحصائية تي لدينا كبيرة بشكل كافٍ بمعنى أن معاملنا المشاهد يبتعد كثيرًا عما يمكن أن تتنبأ به الفرضية المنعدمة، حينها نستطيع رفض الفرضية المنعدمة عند مستوًى ما من الدلالة الإحصائية. مرة أخرى تلك هي نفس عملية الاستدلال الإحصائي الأساسية التي ظللنا نوظفها على مدى الكتاب بأكمله.

كلما قلَّت درجات الحرية (ومن ثم ازدادت أذيال توزيع تي ذي الصلة بها «سُمكًا»)، تَعيَّن أن ترتفع إحصائية تي كي نستطيع أن نرفض الفرضية المنعدمة عند مستوًى ما من مستويات الدلالة. في مثال الانحدار الافتراضي ذلك الذي ذكرناه لتوِّنا، لو كانت لدينا أربع درجات حرية، فإننا نحتاج حينها لأن تكون إحصائية تي على الأقل ٢,١٣ كي نستطيع رفض الفرضية المنعدمة عند مستوى ٠٥. (في تحليل أحادي الذيل).

لكن إذا كان لدينا ٢٠٠٠٠ درجة من درجات الحرية (وهو الأمر الذي يسمح لنا بأن نستخدم التوزيع الطبيعي)، فلا نحتاج إلا لإحصائية تي قدرها ١,٦٥ كي نرفض الفرضية المنعدمة عند مستوى ٠٥. في نفس التحليل الأحادي الذيل.

معادلة الانحدار للوزن

المتغير	المعامل	الخطأ المعياري	إحصائية تي	قيمة p (اختبار ثنائي الذيل)	فاصل ثقة نسبته 95%
الطول	4.4	.2	21.4	.000	4.0 إلى 4.8
العمر	.08	.03	2.2	.026	.01 إلى .2
الجنس	-5.7	1.7	-3.4	.001	-9.0 إلى -2.4
سنوات التحصيل الدراسي	-.7	.2	-3.5	.000	-1.1 إلى -.3
الخُمس الأسفل في ترتيب النشاط البدني	3.7	1.4	2.6	.009	.9 إلى 6.5
متغير وهمي لتلقِّي قسائم الطعام المجاني	5.6	2.1	2.7	.007	1.5 إلى 9.7
سُودُ البشرة من غير الهسبانيين	9.7	1.3	7.2	.000	7.0 إلى 12.3
التداخل	-117				

الفصل الثاني عشر

أخطاء الانحدار الشائعة

بطاقة التحذير الإلزامية

إليك واحدًا من أهم الأمور التي يجب عليك أن تتذكرها بينما تجري بحثك الذي يتضمن تحليل انحدار: حاول ألا تقتل أي أحد. بل يمكنك كذلك أن تلصق ورقة صغيرة على شاشة حاسوبك وتكتب فيها: «لا تقتل الناس ببحثك»؛ وذلك لأن بعضًا من أشد الناس ذكاءً قد خرقوا تلك القاعدة غير عامدين.

فابتداءً من تسعينيات القرن الفائت اتحدت المؤسسة الطبية جمعاء على فكرة أن النساء المتقدمات في السن يجدر بهن أن يتعاطين مكملات الإستروجين، كي تقيهن من مرض القلب وهشاشة العظام وغيرها من الأمراض التي ترتبط بانقطاع الطمث[١]. وبحلول عام ٢٠٠١ كان الإستروجين يوصف طبيًّا لنحو ١٥ مليون سيدة استنادًا للاعتقاد بأنه سيجعلهن أفضل صحة. لماذا؟ لأن الأبحاث في ذلك الوقت – التي كانت تَستخدم نفس المنهجية الأساسية التي شرحناها في الفصل السابق – أفادت بأن تلك استراتيجية طبية سليمة. ولنكون أكثر تحديدًا فقد وجدت دراسة طولية أجريت على ١٢٢٠٠٠ امرأة (دراسة صحة الممرضات) ارتباطًا سلبيًّا بين مكملات الإستروجين والنوبات القلبية. فقد بلغ معدل النوبات القلبية التي أصيبت بها النساء اللاتي تعاطين الإستروجين ثلث نظيره لدى النساء اللاتي لم يتعاطينه. لم يكن القائمون على هذه الدراسة حفنة من المراهقين الذين يستخدمون حاسوب أبيهم كي يتصفحوا المواقع الإباحية ويجروا معادلات الانحدار. وإنما أجريت دراسة صحة الممرضات بواسطة مدرسة هارفارد للطب ومدرسة هارفارد للصحة العامة.

في نفس الفترة قدَّم الأطباء والعلماء نظرية طبية للسبب وراء كون المكملات الهرمونية قد تكون مفيدة لصحة المرأة. مع تقدُّم المرأة في السن تفرز مبايضها كميات أقل من الإستروجين. إذا كان الإستروجين مهمًّا للجسد، فإن تعويض هذا العجز في السن

المتقدمة قد يكون حاميًا لصحة المرأة على المدى الطويل. ومن ثم فقد سمي هذا العلاج: علاج الاستبدال الهرموني. بل إن بعض الباحثين قد اقترحوا أن يتلقى المسنون من الرجال جرعات معززة من الإستروجين[٢]

ومن ثم، وبينما يوصف علاج الاستبدال الهرموني لملايين النساء، كان الإستروجين يخضع لأكثر أنواع التدقيق العلمي صرامة، ألا وهي التجارب السريرية. فتلك التجارب السريرية لا تبحث في مجموعات البيانات الكبيرة – على غرار دراسة صحة الممرضات – عن الارتباطات الإحصائية التي قد تكون أو لا تكون ذات علاقة سببية، وإنما تتألف عوضًا عن ذلك من تجربة ضابطة. تعطى إحدى العينتين علاجًا كعلاج الاستبدال الهرموني وتعطى عينة أخرى علاجًا وهميًّا. أظهرت التجارب السريرية أن النساء اللاتي تعاطين الإستروجين تزايدت لديهن معدلات الإصابة بأمراض القلب والسكتات الدماغية وجلطات الدم وسرطان الثدي، وغيرها من الآثار الصحية السلبية. إن لمكملات الإستروجين بعض الفوائد لكن المخاطر من ورائه أكثر بكثير. ابتداءً من عام ٢٠٠٢ نُصح الأطباء بألَّا يصفوا الإستروجين لمريضاتهم المتقدمات في العمر. طرحت مجلة النيويورك تايمز سؤالًا حساسًا، وإن كان مهمًّا من الناحية الاجتماعية، ألا وهو: كم من النساء توفين قبل أوانهن أو عانَين السكتات الدماغية أو سرطان الثدي لأنهن تناولن أقراص الدواء التي وصفها لهن أطباؤهن كي تحافظ على صحتهن؟

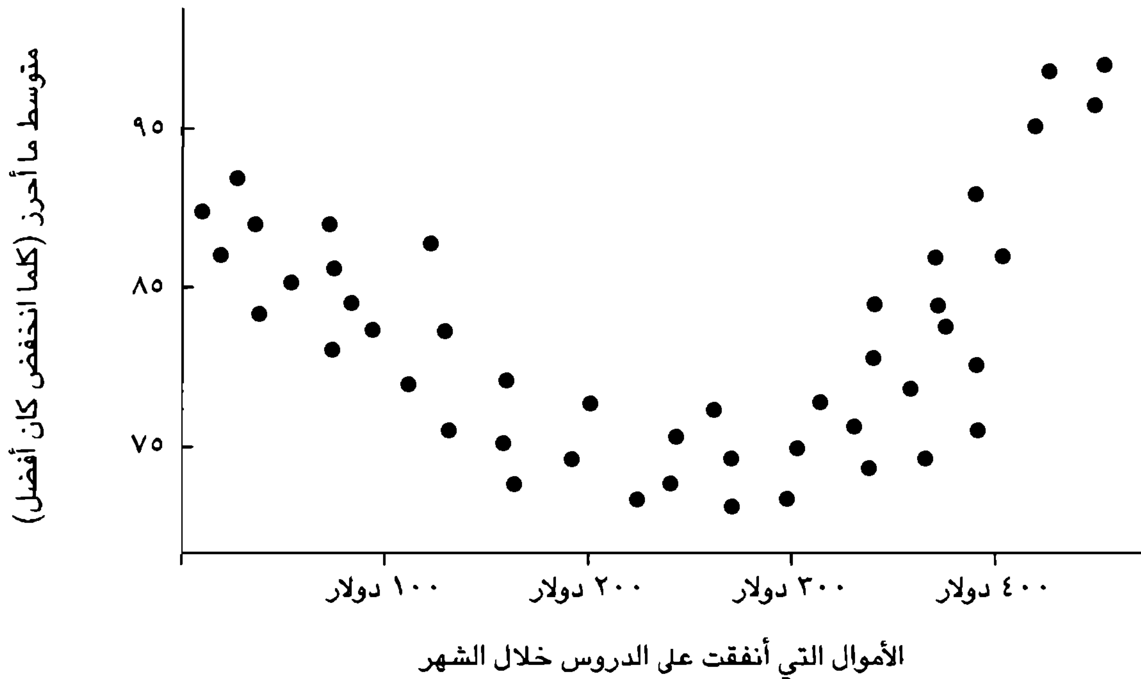
الإجابة: «التقدير المعقول لهذا العدد قد يكون عشرات الآلاف»[٣]

إن تحليل الانحدار هو بمثابة القنبلة الهيدروجينية في الترسانة الإحصائية. فأي شخص يمتلك حاسوبًا شخصيًّا ومجموعة كبيرة من البيانات، يمكن أن يصير باحثًا وهو داخل منزله أو حجيرته الصغيرة في مقر عمله. ما الذي يمكن أن يسير بشكل خطأ؟ كل شيء يمكن أن يسير بالشكل الخطأ. يقدم تحليل الانحدار إجابات محددة لأسئلة معقدة. هذه الأجوبة قد تكون دقيقة أو لا تكون. وإذا أجري تحليل الانحدار بأيدي الأشخاص الخطأ، فسوف يخرج بنتائج مضللة أو خطأ جدًّا. وكما أوضح مثال الإستروجين، حتى ولو أجري تحليل الانحدار بأيدي الأشخاص المناسبين، فقد تدفعنا تلك الأداة الإحصائية القوية بسرعة خطيرة نحو الاتجاه الخطأ. سوف يقوم هذا الفصل بشرح أكثر «أخطاء» الانحدار شيوعًا. ولقد وضعت كلمة «أخطاء» بين علامتَي تنصيص لأنه، كما في جميع أنواع التحليل الإحصائي، يمكن للأشخاص البارعين أن يستغلوا هذه النقاط المنهجية عمدًا لتحقيق أغراض مشينة.

إليك قائمة «السبعة الكبار» من بين أكثر ضروب إساءة استعمال تلك الأداة المذهلة شيوعًا.

استخدام الانحدار في تحليل علاقة غير خطية*. هل قرأت يومًا بطاقة التحذير الملصقة على مجفف الشعر – ذلك الجزء الذي يقول: لا تستخدمه في حوض الاستحمام – ففكرت في نفسك «من ذا الأحمق الذي يستخدم مجفف شعر وهو في حوض الاستحمام؟!» إنه جهاز كهربائي، لا يستخدم المرء جهازًا كهربائيًّا قرب الماء؛ فهو غير مصمم لذلك. ولو كان لتحليل الانحدار بطاقة تحذيرية مماثلة لكُتب عليها «لا تستخدمه إذا لم يكن هناك ارتباط خطي بين المتغيرات التي تحللها». تذكر أن معامل الانحدار يصف ميل «خط أفضل مطابقة» للبيانات، فالخط غير المستقيم سيميل في مواضع مختلفة. ومثالًا على هذا لنتأمل تلك العلاقة الفرضية التالية بين عدد دروس الجولف التي أتلقاها خلال شهر واحد (متغير مفسر)، ومتوسط ما أُحرزه في جولة من جولات المباريات ذات الثماني عشرة حفرة خلال نفس هذا الشهر (المتغير الاعتمادي). كما ترى من توزيع الانتشار هذا، فليس هناك من علاقة خطية متواصلة.

تأثير دروس الجولف على الأهداف المحرزة



هنالك نمط لكنه ليس من السهل أن يوصف بخط وحيد مستقيم. فدروس الجولف القليلة الأولى قد بدا أنها هبطت بنقاطي بشكل سريع. هناك ارتباط سلبي بين الدروس

* هنالك مناهج أكثر تعقيدًا يمكن أن تستخدم لتكييف تحليل الانحدار مع البيانات غير الخطية. لكنك قبل أن تستخدم تلك الأدوات لا بد تفهم لماذا سيؤدي استخدامك لمنهج المربعات الصغرى العادية مع البيانات غير الخطية إلى أن تخرج بنتائج غير ذات معنًى.

وبين النقاط التي أحرزتها في هذه المدة، الميل هنا سلبي. فالمزيد من الدروس ينتج عنها نقاط أقل (وهو أمر جيد في الجولف)*.

لكنني عندما أصل للنقطة التي أنفق فيها ما بين ٢٠٠ و٣٠٠ دولار شهريًّا على تلك الدروس، لا يبدو أن للدروس تأثيرًا على الإطلاق. ليس هناك ارتباط واضح خلال هذه المدة بين التعليم الإضافي والنقاط التي أحرزها، هنا يكون الميل صفرًا.

وأخيرًا تصير الدروس ذات نتائج عكسية. فما أن أُصبح أنفق ٣٠٠ دولار شهريًّا على تلك الدروس، حتى تصير تلك الدروس التصاعدية ترتبط ارتباطًا إيجابيًّا بإحراز نقاط أكثر، هنا يكون الميل إيجابيًّا خلال هذه المدة. (سوف أناقش تلك الاحتمالات المختلفة لأن يكون سوء مستواي في الجولف هو الذي دفعني لآخذ تلك الدروس، لا العكس، في مكان لاحق من هذا الفصل).

النقطة الأهم هنا هي أننا لا نستطيع تلخيص العلاقة بين الدروس والنقاط المحرزة بدقة من خلال معامل واحد فقط. أفضل تفسير للنمط الذي وصفناه لتوِّنا هو أن دروس الجولف لها علاقات خطية مختلفة مع النقاط التي أحرزتها. يمكنك أنت أن ترى هذا، لكن حزمة البرامج الإحصائية لن تراه. فلو غذَّيتَ معادلة انحدار بهذه البيانات، فسوف يَخرج لك الحاسوب بمعامل واحد. وهذا المعامل لن يعكس بدقةٍ العلاقةَ الحقيقية مع مختلف المتغيرات التي تهمنا. وستصير النتائج التي تحصل عليها كاستخدام مجفف شعر داخل حوض الاستحمام.

إن المستهدف من تحليل الانحدار هو أن يستخدم عندما يتسنى التعبير عن العلاقة بين المتغيرات من خلال معادلة خطية. (هناك طرق يتم من خلالها تحويل العلاقة غير الخطية إلى أخرى خطية، كأن نأخذ الخوارزمية كمتغير، لكننا لن نفعل ذلك هنا). أي كتاب أكاديمي أو برنامج تدريبي متقدم في الإحصاء، سوف يعلمك خطوة بخطوة الفرضياتِ الأساسيةَ الأخرى التي ينطوي عليها تحليل الانحدار. وكما في غيرها من الأدوات الأخرى، كلما انحرفت الصور المتقدمة منها عن غايتها الأساسية، صارت أقل فعالية، بل وحتى أكثر خطورة.

الارتباط لا يساوي السببية. يقوم تحليل الانحدار فقط بعرض الارتباط بين متغيرين. وكما ذكرت سابقًا لا نستطيع أن نثبت بالإحصاء وحده أن تغيرًا في واحد من المتغيرات يسبب التغير في الآخر. بل إنه لو أجريت معادلة انحدار بشكل فيه إهمال، فإن هذا يمكن أن يُنتج رابطًا إحصائيًّا ذا دلالة مهمة بين متغيرين لا علاقة لأحدهما بالآخر. لنفترض أننا نفتش عن أسباب محتملة لارتفاع معدلات التوحد في الولايات المتحدة خلال

* تشير النقاط المحرزة هنا إلى عدد الضربات التي ضربها اللاعب لإدخال الكرة في الحفرة. وفي لعبة الجولف كلما قل عدد تلك الضربات، كان ذلك أفضل (المترجم).

العقدين الأخيرين. سيكون متغيرنا الاعتمادي – المخرج الذي ننشد تفسيره – مقياسًا ما لمعدل التوحد خلال عام واحد، كعدد الحالات المشخصة من بين كل ١٠٠٠ طفل من عمر معين. لو ضمَّنَّا التحليل الدخل السنوي في الصين كمتغير مفسر، فمِن شبه المؤكد أننا سوف نجد رابطًا ذا دلالة إيجابية وإحصائية بين ارتفاع الدخول في الصين ومعدلات الإصابة بالتوحد في الولايات المتحدة على مدى العشرين عامًا الفائتة.

لماذا؟ لأنه كلاهما شهدا ارتفاعات حادة على مدى نفس الفترة. لكنني أشك كثيرًا في أنه إذا ما حدث كساد حاد في الصين، فسيؤدي ذلك إلى تقليل معدلات التوحد في الولايات المتحدة الأمريكية. (كي أكون منصفًا، لو لاحظت علاقة قوية بين نمو اقتصادي سريع حادث في الصين ومعدلات التوحد في الصين وحدها، لشرعت أفتش عن عامل بيئي ما ذي صلة بالنمو الاقتصادي، كالتلوث الصناعي، والذي قد يفسر ذلك الارتباط).

ذلك النوع من الارتباط الزائف بين متغيرين، والذي وضحته لتوِّي، ما هو إلا مثال واحد على ظاهرة أكثر عمومية تُعرف باسم السببية الزائفة. هنالك أكثر من طريقة يمكن أن يفسر بها الرابط بين (أ) و(ب) بطريقة خطأ.

السببية العكسية. الارتباط الإحصائي بين (أ) و(ب) لا يثبت أن (أ) يسبب (ب)، وإنما من الممكن جدًّا أن يكون (ب) هو الذي يسبب (أ). كنت قد ألمحت لهذه الاحتمالات سابقًا في مثال دروس الجولف. لنفترض أنني عندما أبني نموذجًا معقدًا لتفسير ما أحرزه من نقاط في لعبة الجولف، سيصير متغير دروس الجولف مرتبطًا ارتباطًا متواصلًا مع أسوأ النتائج. كلما تلقيت دروسًا أكثر، ضربت الكرة بشكل أسوأ. واحد من تفسيرات ذلك أن مدرسي للجولف كان مدربًا سيئًا جدًّا جدًّا. أما التفسير الأكثر عقلانية هو أنني لا أتلقى تلك الدروس إلا لأنني سيء في اللعب، فممارسة الجولف بشكل سيء هو ما يسبب المزيد من الدروس لا العكس. (هناك بعض التصحيحات المنهجية لمشكلة من هذا النوع. فمثلًا أستطيع أن أُضمِّن دروس الجولف التي تلقيتها في شهر ما كمتغير مفسر لنتائج الجولف في الشهر الذي يليه).

وكما نوَّهنا في محل سابق من هذا الفصل، قد تسير السببية في كلا الاتجاهين. لنفترض أنك تقوم ببحث يُظهر أن الولايات التي تنفق المزيد من المال على التعليم ذي الاثني عشر صفًّا، تتمتع بمعدلات نمو تتفوق على مثيلاتها لدى الولايات التي تنفق أقل على نفس النوع من التعليم. لن يمنحنا أيُّ ارتباط إيجابي ذي دلالة بين هذين المتغيرين أيَّ نتيجة عن الاتجاه الذي تسير فيه تلك العلاقة. الاستثمار في التعليم ذي الاثني عشر صفًّا قد يسبب نموًّا اقتصاديًّا. لكن في المقابل تستطيع الولايات ذات الاقتصادات القوية تحمُّل تكلفة إنفاق المزيد من الأموال على التعليم ذي الاثني عشر صفًّا؛ لذا فإن الاقتصاد القوي قد يسبب المزيد من الإنفاق على التعليم، أو قد يعطي الإنفاق على التعليم دفعة قوية للنمو

الاقتصادي، وهو الذي يتيح المزيد من الإنفاق على التعليم – فإن السببية يمكن أن تسير في كلا الاتجاهين.

المقصد هنا هو أنه لا يجدر بنا أن نستخدم المتغيرات المفسرة التي يمكن أن تكون قد تأثرت بالمخرجات التي نحاول تفسيرها، وإلا خرجت النتائج متشابكة لا أمل في حلها. فمثلًا لن يكون من المناسب أن نستخدم معدل البطالة في معادلة انحدار تفسر الناتج الإجمالي المحلي، لأنه من الواضح أن البطالة تتأثر بمعدل الناتج المحلي. أو، لو فكرت بالأمر بطريقة أخرى، لو خلص تحليل الانحدار إلى أن خفض البطالة سوف يؤدي إلى تعزيز إجمالي الناتج المحلي، فهي نتيجة فارغة لا معنى لها، لأن خفض البطالة يحتاج لتعزيز الناتج المحلي.

يجدر بنا أن نمتلك سببًا للاعتقاد أن متغيراتنا المفسرة تؤثر على المتغير الاعتمادي، لا العكس.

انحياز المتغير المحذوف. يجدر بك أن تتشكك إذا ما رأيت عنوانًا عريضًا يقول «ممارسو الجولف أكثر عرضة لأمراض القلب والسرطان والتهاب المفاصل». لن أندهش إذا ما كان ممارسو الجولف يصابون بهذه الأمراض بمعدلات أعلى من مثيلاتها لدى من لا يمارسون تلك اللعبة، وإنني كذلك أتوقع أن الجولف مفيد للصحة لأنه ينطوي على اختلاط اجتماعي وتدريب خفيف. كيف لي أن أوفق بين هاتين المقولتين؟ هو أمر سهل جدًّا. فأي دراسة تحاول قياس تأثيرات لعب الجولف على الصحة لا بد لها أن تتخذ العمر متغيرًا ضابطًا. فبشكل عام يزيد الناس من لعبهم للجولف مع تقدُّمهم في السن، وبالأخص بعد أن يتقاعدوا. لذا فإن أي تحليل يغفل عن العمر كمتغير مفسر، سوف تغيب عنه حقيقة أن لاعبي الجولف هم بشكل عام أكبر عمرًا ممن لا يلعبونها. فالجولف لا يقتل الناس، وإنما كبر السن يقتلهم، وهم يستمتعون باللعبة بينما يقتلهم. وإنني لأظن أن العمر عندما يدخل في تحليل الانحدار كمتغير ضابط، سيسبب الحصول على مخرج مختلف. وقد يكون الجولف بين المتماثلين في الأعمار واقيًا محدودًا من أمراض خطيرة. ذلك فارق كبير جدًّا.

في هذا المثال يكون العمر «متغيرًا محذوفًا» مهمًّا. عندما نستبعد العمر من معادلة انحدار تفسر مرض القلب أو غيره من المخرجات الصحية السلبية، حينها يؤدي متغير «لعب الجولف» دورين تفسيريين، لا واحدًا. فهو ينبئنا بتأثير لعب الجولف على مرض القلب، وينبئنا بتأثير التقدم في العمر على مرض القلب (بما أن ممارسي الجولف يميلون لأن يكونوا أكبر سنًّا من بقية مجتمعهم). نستطيع أن نقول بمصطلحات الإحصاء إن متغير الجولف «يلتقط» تأثير العمر. المشكلة هنا هي أن هذين التأثيرين مختلطان. مما سيجعل نتائجنا في أفضل الأحوال فوضوية عشوائية، وفي أسوأ الأحوال سيجلعنا نفترض أن الجولف ضار بالصحة، بينما النقيض هو الحق على الأرجح.

ستصير نتائج الانحدار مضللة وغير دقيقة إذا ما استَبعدت معادلةُ الانحدار متغيرًا مفسرًا مهمًّا، وخاصة إذا ما «التقطت» بقية المتغيرات ذلك التأثير. لنفترض أننا نحاول تفسير جودة مدرسة ما؟ إنه مخرج مهم يجدر بنا أن نفهمه: ما الذي يجعل المدارس جيدة؟ متغيرنا الاعتمادي – مقياس الجودة الذي يقدر كميًّا – سيكون على الأرجح نتائج الاختبارات. إنه من شبه المؤكد أننا سوف نفحص الإنفاق المدرسي، بصفته أحد المتغيرات المفسرة، طلبًا لتقدير العلاقة بين الإنفاق ونتائج الاختبارات تقديرًا كميًّا. هل تحرز المدارس التي تنفق المزيد من الأموال نتائج أفضل؟ لو كان الإنفاق المدرسي هو المتغير المفسر الوحيد، فلا شكَّ لدي في أننا سوف نجد علاقة كبيرة دالة إحصائيًّا بين الإنفاق ونتائج الاختبارات. لكن هذه النتيجة وما تنطوي عليه من إيحاءات بأننا نستطيع أن نحسن مدارسنا بالمزيد من الإنفاق هي نتيجة بها الكثير من العطب.

هناك الكثير من المتغيرات المحذوفة التي يحتمل أن تكون ذات دلالة، لكن المتغير الحاسم هنا هو مستوى تعليم الوالدين. فالأسر الأعلى تعليمًا عادة ما تقطن مناطق أكثر ثراءً تنفق الكثير من الأموال على مدارسها، وهذه الأسر عادة ما يبلي أبناؤها بلاءً حسنًا في الاختبارات (وأبناء الأسر الفقيرة عادة ما يلاقون صعوبات في الاختبارات). لو لم يتوافر لدينا مقياس ما للوضع الاجتماعي الاقتصادي للتلاميذ كمتغير الضبط، فسوف تُظهر نتائج الانحدار غالبًا ارتباطًا إيجابيًّا كبيرًا بين الإنفاق المدرسي ونتائج الاختبارات – في حين أن تلك النتائج في الواقع قد لا تكون إلا نتاجًا لطبيعة التلاميذ الذين يدلفون من باب المدرسة لا للأموال التي تنفق على مبناها.

أذكر أن أحد أساتذتي في الجامعة قد قال إن نتائج اختبارات SAT ترتبط بعدد السيارات التي تمتلكها الأسرة. وكان بهذا يلمِّح إلى أن اختبارات SAT هي وسيلة غير منصفة وغير مناسبة تستخدم للقبول في الجامعات. تلك الاختبارات بها من العيوب ما بها، لكن الارتباط بين نتائج الاختبارات وسيارات الأسرة ليس هو أكثر ما يهمني. فلا يشغلني كثيرًا كيف أن الأسر الثرية تُدخل أبناءها إلى الجامعات عن طريق شرائها ثلاث سيارات إضافية. فعدد السيارات الرابضة في مرآب منزل الأسرة ما هو إلا مؤشر على دخل الأسرة، ومستوى تعليمها، وغيرها من مقاييس الوضع الاجتماعي الاقتصادي. وحقيقة أن الأطفال الأثرياء يُبلون أفضل من الفقراء في اختبارات SAT ليست بالخبر الجديد. (كما أشرنا من قبل، يبلغ متوسط نتائج اختبار القراءة النقدية في SAT لأبناء الأسر التي يتعدى دخلها ٢٠٠٠٠٠ دولار أمريكي، ١٣٤ نقطة أعلى من متوسط نتائج اختبار أطفال الأسر التي يقل دخلها عن ٢٠٠٠٠ دولار أمريكي)[٤]. لكن أكبر همِّنا هنا يجب أن يكون إذا ما كان اختبار SAT «سهل الفهم» أو لا. كم يستطيع التلاميذ أن يحسِّنوا من نتائجهم عن طريق أخذ دروس خاصة؟ بالطبع فإن الأسر الثرية أكثر قدرة على إرسال أبنائها إلى دروس

خاصة لإعدادهم لاختبارات SAT. إذا ما حدث أي تحسن سببي بين هذه الصفوف الخاصة ونتائج اختبارات SAT، فسوف تأتي في صالح أبناء الأسر الثرية، في مقابل أبناء الأسر المحرومة من ذوي القدرات المساوية (الذين يفترض أنهم كانت نتائجهم سترتفع لو تلقوا تلك الدروس الخاصة، لكنهم لم تتح لهم الفرصة لذلك).

المتغيرات المفسرة شديدة الارتباط بعضها ببعض (التداخل الخطي). لو تضمنت معادلة الانحدار اثنين أو أكثر من المتغيرات المفسرة التي ترتبط بعضها ببعض ارتباطًا شديدًا، فلن يستطيع التحليل بالضرورة أن يظهر العلاقة الحقيقية بين كل من تلك المتغيرات والمخرج الذي نحاول تفسيره. سيتضح ذلك أكثر من خلال مثال. لنفترض أننا نحاول قياس تأثير المخدرات الممنوعة على نتائج اختبارات SAT. وعلى وجه التحديد، فإن لدينا بيانات تخبرنا بما إذا كان المشاركون في دراستنا قد سبق لهم وتعاطوا الكوكايين، أو إذا ما سبق لهم وتعاطوا الهروين. (من المفترض أن يكون لدينا الكثير من متغيرات الضبط الأخرى كذلك). ما تأثير الكوكايين على نتائج اختبارات SAT، مع تثبيت العوامل الأخرى، ومن بينها تعاطي الهروين؟ وما تأثير تعاطي الهروين على نتائج اختبارات SAT، مع تثبيت تعاطي الكوكايين وغيره من العوامل؟

قد لا تستطيع معاملات تعاطي الهروين والكوكايين أن تخبرنا بذلك. فالتحدي المنهجي يكمن هنا في كون من تعاطوا الهروين قد قاموا على الأرجح بتعاطي الكوكايين. إذا وضعنا المتغيرين كليهما في المعادلة، فسيصير لدينا أفراد قليلون جدًّا ممن تعاطوا واحدًا فقط من المخدرين دون الآخر، مما يترك لنا قليلًا جدًّا من التغاير في البيانات التي سنحسب بها تأثيراتهما المستقلة. لنعُد للحظة إلى الصورة الذهنية التي استخدمناها لتفسير تحليل الانحدار في الفصل السابق. سوف نقوم بتقسيم عينة بياناتنا إلى «غرف» مختلفة تتطابق فيها المشاهدات إلا في متغير واحد، مما يسمح لنا بأن نعزل تأثير ذلك المتغير بينما نتحكم في العوامل الأخرى التي يحتمل أن تكون مربكة. قد نجد لدينا ٦٩٢ فردًا من المشاركين في العينة سبق لهم وأن تعاطوا الهروين والكوكايين كليهما. لكن قد يكون لدينا ٣ أفراد تعاطوا الكوكايين دون الهروين، وفردان تعاطيا الهروين دون الكوكايين. أيُّ استدلال عن التأثير المستقل لمخدر واحد دون الآخر سيقوم على هذه العينات الصغيرة.

من غير المحتمل أن نحصل على معاملات ذات معنًى، سواء من متغير الكوكايين أو من متغير الهروين، بل إننا قد نشوش على تلك العلاقة الأكبر والأهم، والقائمة بين نتائج اختبارات SAT وتعاطي واحدٍ من هذين المخدرين. عندما يكون هناك متغيران مفسران مرتبطان أحدهما بالآخر بشدة، يستخدم الباحثون عادة أحدهما في معادلة الانحدار، أو قد يقومون بخلق نوعٍ من أنواع المتغير المركب على غرار «تعاطى إما الهروين أو الكوكايين». فمثلًا عندما يريد الباحثون أن يضبطوا خلفية الطالب الاجتماعية الاقتصادية في مجملها،

فقد يُضمنون التحليل متغيرات لكلٍّ من «مستوى تعليم الأم» و«مستوى تعليم الأب»، لأن هذا التضمين قد يقدم لهم فكرة عن الخلفية التعليمية للأسرة كلها. لكن إذا كان هدف التحليل هو عزل تأثير مستوى تعليم واحد من الوالدين، يصير وضع كلا المتغيرين في المعادلة في الغالب مشوِّشًا للمسألة لا موضِّحًا لها. إن الارتباط بين مستوى تعليم الزوج ومستوى تعليم الزوجة عالٍ جدًّا، لدرجة أننا لا نستطيع أن نعتمد على تحليل الانحدار في الحصول على المعاملات التي تقوم بعزلٍ ذي معنًى لتأثير المستوى التعليمي لأحد الوالدين (كما هو من الصعب الفصل بين تأثير تعاطي الكوكايين وتأثير تعاطي الهروين).

استقراء ما وراء البيانات. إن تحليل الانحدار مثل جميع صور الاستدلال الإحصائي، مصمم ليقدِّم لنا فكرة بصيرة عن العالم من حولنا. إننا ننشد أنماطًا تصح على المجتمع الحقيقي الأكبر. لكن نتائجنا لا تصح إلا على المجتمع الذي يشابه العينة التي نجري عليها التحليل. في الفصل السابق قمت بخلق معادلة انحدار للتنبؤ بالوزن استنادًا إلى عدد المتغيرات الاعتمادية. كان الـ R^2 للنموذج النهائي ٢٩. وهو ما يعني أنه قد أجاد تفسير التغاير في الوزن لعينة كبيرة مؤلفة من أفراد – تصادف أنهم جميعًا بالغون.

إذن ماذا سيحدث لو استخدمنا معادلتنا للانحدار في التنبؤ بالوزن الأرجح لطفل حديث الولادة؟ لنجرب. بلغت ابنتي إحدى وعشرين بوصة طولًا وقت أن ولدت. سنقول إن عمرها وقت الولادة كان صفرًا، ولم تتلقَّ أي تعليم أو تدريب. كانت أنثى بيضاء اللون. لو حسبنا معادلة الانحدار بناءً على بيانات دراسة تغير الحيوات، لتنبأت تلك المعادلة بأن وزنها حين الولادة سيكون ١٩,٦ رطلًا سالبًا (كان وزنها الحقيقي ثمانية أرطال ونصفًا).

قام واضعوا إحدى دراسات وايت هول، والتي أشرنا إليها في الفصل السابق، بتصريح صادم عن خلاصتهم الضيقة: «إن ضَعف التحكم في بيئة العمل يرتبط بازدياد خطورة الإصابة بمرض الشريان التاجي مستقبلًا بين الرجال والنساء الذين يعملون لصالح الحكومة»[٥] (مع التركيز على تلك العبارة الأخيرة).

التنقيب عن البيانات (متغيرات أكثر مما ينبغي). لو كان حذف المتغيرات المهمة هو من المشكلات المحتملة، فمن المفترض أن تكون إضافة متغيرات مفسرة كثيرة إلى معادلة الانحدار، قدر ما نستطيع، هي الحل. لا.

قد ترتبك نتائجك إذا ما ضمَّنْتَ تحليلك متغيرات أكثر مما ينبغي، وخاصة لو كانت متغيرات مفسرة خارجية لا مبرر نظري لها. فمثلًا لا ينبغي للمرء أن يصمم استراتيجية بحثية بناءً على الفرضية التالية: بما أننا لا نعرف ما يسبب التوحد، فيجدر بنا أن نضع في معادلة الانحدار أكبر قدر ممكن من المتغيرات التي يحتمل أن تكون مفسرة، فقط، لنرى أيها ستتضح له دلالة إحصائية، بعدها قد نتحصل على بعض الإجابات. لو وضعنا قدرًا

كافيًا من المتغيرات غير المرغوب فيها في معادلة الانحدار، فلا بد أن واحدًا منها سوف يلاقي حد الدلالة الإحصائية عن طريق الصدفة. والأكثر خطورة أن تلك المتغيرات غير المرغوب فيها ليس من السهل التعرف على طبيعتها تلك. يستطيع الباحثون المهرة دومًا أن يبنوا نظرية، بعد فوات الأوان، عن السبب في كون متغيرٍ غريبٍ، ما هو في الحقيقة إلا محض هراء، يتضح بعدها أنه ذو دلالة إحصائية.

لإثبات هذه النقطة، عادة ما أجري تمرين رمي العملة، الذي شرحته خلال مناقشة الاحتمالات. سوف أطلب من صف به أربعين طالبًا أو نحوهم أن يرمي كلٌّ منهم عملة معدنية في الهواء. وكل من تستقر عملته كاشفة وجه الذيل سيتم استبعاده، أما البقية فسيرمون العملة مرة أخرى. وفي الجولة الثانية مرة أخرى سوف يتم استبعاد من تستقر عملاتهم كاشفة وجه الذيل. أستمر في جولات رمي العملة حتى يكون أحد الطلبة قد رمى عملته خمس أو ست مرات متتالية استقرت فيها عملته كاشفة وجه الرأس. لعلك تذكر بعضًا من تلك الأسئلة السخيفة التي ذكرناها سابقًا، والتي تلي تلك العملية على غرار: «ما سرك؟ هل يكمن السر في معصم المرء؟ هل لك أن تعلمنا أن نرمي العملة فتستقر كاشفة وجه الرأس كل مرة؟ لربما كان السر في قميص هارفارد ذلك الذي ترتديه».

من الواضح أن ذلك التسلسل المتتالي من سقوط العملة كاشفة وجه الرأس ما هو إلا من قبيل الحظ، وقد رآه الطلبة كلهم يحدث. لكن رغم ذلك، ليس هذا بالضرورة هو الشكل الذي ستفسر به تلك النتائج في سياق علمي. إن احتمالات إلقاء العملة لتستقر كاشفة وجه الرأس خمس مرات متتالية هي ١/٣٢ أو ٠٫٠٣. وهو القدر الذي يدنو بهامش مريح عن حد ٠٫٠٥ الذي عادة ما نستخدمه لرفض الفرضية المنعدمة. فرضيتنا المنعدمة في هذه الحالة هي أن الطالب لا يمتلك أي موهبة خاصة في رمي العملات. وسلسلة الرءوس المحظوظة تلك (والتي يلزم أن تحدث لطالب واحد على الأقل إذا ما بدأنا بمجموعة كبيرة) تسمح لنا بأن نرفض الفرضية المنعدمة وأن نتبنى الفرضية البديلة، والتي تقول بأن هذا الطالب يمتلك قدرة خاصة على رمي العملات فتسقط كاشفة وجه الرأس. وبعد أن أنجز هذا العمل الخارق، نستطيع أن ندرُسه باحثين عن أدلة على نجاحه في رمي العملات، وطريقة رميته وتدريبه الرياضي وتركيزه الخارق بينما العملة تطير في الهواء، وهلم جرًّا. وما كل هذا إلا محض هراء.

تلك ظاهرة قادرة على أن تفسد حتى الأبحاث الشرعية. إن المفهوم المتفق عليه هو أن ترفض الفرضية المنعدمة عندما نلاحظ أمرًا قد يَحدث عن طريق الصدفة مرة من كل عشرين مرة أو أقل إذا ما صدقت الفرضية المنعدمة. بالطبع لو أجرينا عشرين دراسة أو لو أدخلنا عشرين متغيرًا غير مرغوب فيه في معادلة انحدار واحدة، فسوف نحصل على ما متوسطه نتيجة واحدة زائفة ذات دلالة إحصائية. عبرت مجلة النيويورك تايمز عن هذه

المعضلة ببراعة من خلال اقتباس عن ريتشارد بيتو الإحصائي الطبي وخبير الأوبئة قال فيه: «إن علم الأوبئة علم جميل يزودنا برؤًى مهمة عن الحياة البشرية وعن الموت، لكن الكثير من الهراء ينشر في إطاره»[٦]

وحتى نتائج التجارب السريرية، والتي هي في الغالب تجارب مجراة عشوائية، ومن ثم فهي المعيار الذهبي للبحث الطبي، يجدر بنا أن نستعرضها بتشكك. في عام ٢٠١١ نشرت صحيفة وول ستريت جورنال تحقيقًا في صفحتها الأولى يتناول ما وصَفته بواحد من «الأسرار الصغيرة القذرة» في البحث الطبي: «معظم النتائج، ومن بينها تلك التي تظهر في أشهر الدوريات العلمية وأعرقها من التي يراجعها علماء نظراء للناشرين، لا يمكن استنساخها»[٧] (الدوريات التي يراجعها نظراء هي مطبوعات تراجع الدراسات والمقالات المنشورة بها من حيث الصحة المنهجية من قِبل خبراء في نفس المجال قبل أن يصرح بنشرها، وتعتبر مثل تلك المطبوعات حارس بوابة البحث الأكاديمي). أحد أسباب حدوث هذا «السر الصغير القذر» هو انحياز النشر الإيجابي، الذي وصفناه في الفصل السابع. لو أَوْلى الباحثون والدوريات الطبية الاهتمام إلى النتائج الإيجابية وتجاهلوا النتائج السلبية، فلربما نجحوا في نشر الدراسة الوحيدة التي تجد عقَّارًا فعَّالًا وتجاهلوا العشرين الأخريات غير ذوات الفعالية. وكذلك فإن بعض التجارب السريرية لها عينات صغيرة (كما في حالات الأمراض النادرة)، وهو الأمر الذي يعظِّم فرص أن يَلقى التغاير العشوائي في البيانات من الاهتمام أكثر مما يستحق. وفوق هذا وذاك، قد يتحيز الباحثون، عن وعي أو دون وعي؛ إما بسبب اعتقاد قوي اعتنقوه سابقًا، أو لأنهم إذا ما خرجوا بنتيجة إيجابية فسيأتي هذا بمردود حسن على مسيرتهم المهنية. (فلا أحد أبدًا يغدو أكثر ثراءً أو شهرة لأنه أثبت أن عقَّارًا ما لا يعالج السرطان).

لأجل هذه الأسباب كلها اكتُشف أن قدرًا هائلًا من الأبحاث التي أجراها خبراء خطأ. قام جون يوانيديس الطبيب اليوناني والخبير بعلم الأوبئة بفحص تسعة وأربعين دراسة نشرت في ثلاث من الدوريات الطبية البارزة[٨]. وكل واحدة من تلك الدراسات قد اقتبس منها في المراجع الطبية ألف مرة على الأقل، لكن ثلث هذه الأبحاث قد فنَّدَته أبحاث لاحقة. (فمثلًا كانت بعض الدراسات التي فحصها تروِّج لعلاج استبدال الإستروجين). قدَّر الدكتور يوانيديس أن نحو نصف الأبحاث العلمية المنشورة لا بد وأن يتضح خطؤها في النهاية[٩]. نشر بحثه في جورنال أوف أمريكان ميديكال أسوسياشن، وهي إحدى الدوريات التي نشرت بها المقالات التي فحصها. وهو الأمر الذي يخلق مفارقة تدير الرأس: فلو كان بحث الدكتور يوانيديس صحيحًا، فهناك احتمال قوي أن يكون بحثه خطأً.

ما زال تحليل الانحدار أداة إحصائية رائعة. (نعم، لربما كان وصفي إياه في الفصل السابق بأنه «الإكسير المعجزة» مبالغًا بعض الشيء). يمكننا تحليل الانحدار من إيجاد

الأنماط الأساسية في داخل مجموعات البيانات الكبيرة، وتلك الأنماط عادةً ما تكون مفتاحًا للأبحاث المهمة، سواء في الطب أو في العلوم الاجتماعية. والإحصاء يقدم لنا معايير موضوعية لتقييم تلك الأنماط. وعندما يستخدم تحليل الانحدار بشكل صحيح، يصير جزءًا مهمًّا من المنهج العلمي. انظر لهذا الفصل كبطاقة تحذير إلزامية.

وكل تلك التحذيرات المتنوعة والمحددة المكتوبة على تلك البطاقة يمكن أن تلخص في فكرتين. أولاهما أن تصميم معادلة انحدار جيدة – بأن نحدد أي المتغيرات ينبغي أن تفحص ومن أين يجب أن تأتي البيانات – هو أمر أكثر أهمية من الحسابات الإحصائية التي ينطوي ذلك الأمر عليها. تعرف هذه العملية باسم تقدير المعادلة أو تحديد معادلة انحدار جيدة. وأفضل الباحثين هم القادرون على التفكير بمنطق في أي المتغيرات التي ينبغي أن تضمَّن في معادلة الانحدار، وأيها غائب، وكيف ينبغي أن تفسر النتائج النهائية.

وثانيتهما أن تحليل الانحدار كغيره من غالبية أساليب الاستدلال الإحصائي لا يبني سوى حالة ظرفية. فالارتباط بين اثنين من المتغيرات هو أشبه ببصمة إصبع في مسرح جريمة، يشير لنا ناحية الاتجاه الصحيح لكنه نادرًا ما يكفي لإدانة المجرم. (وأحيانًا تكون بصمة الإصبع تلك الموجودة في مسرح الجريمة لا تخص المجرم). يحتاج أي تحليل انحدار إلى دعم نظري: لماذا هي تلك المتغيرات المفسرة في المعادلة؟ أي ظواهر المناهج الأخرى قد تفسر تلك النتائج الملاحظة؟ فمثلًا لماذا نظن أن انتعال أحذية قرمزية اللون سوف يحسن من أدائنا في قسم الرياضيات في اختبارات SAT، أو أن تناوُل الفشار يمكن أن يساعد في الوقاية من سرطان البروستاتا؟ يجب أن تكون النتائج متماثلة أو على الأقل متسقة مع نتائج أخرى.

فحتى الإكسير المعجزة لن يؤدي غرضه إذا لم يُتناول كما تقول التعليمات.

الفصل الثالث عشر

تقييم البرنامج

هل سيغير ارتيادك لهارفارد من حياتك؟

لا يعد الباحثون العباقرة في العلوم الاجتماعية عباقرة لمجرد أنهم يستطيعون إجراء الحسابات المعقدة في أذهانهم، أو أنهم يربحون مالًا في برنامج المسابقات Jeopardy أكثر من زملائهم الباحثين الأقل عبقرية (رغم أن كلا هذين العملين قد يكون صحيحًا). الباحثون العباقرة – أولئك الذين يغيرون معرفتنا بالعالم بطريقة محمودة – هم عادة الأفراد أو الفِرَق الذين يتوصلون لطرق مبدعة لإجراء التجارب «الضابطة». ولكي نقيس تأثير أي معالجة أو أي تداخل، نحتاج لشيء نقيسه في مقابله. كيف سيؤثر ارتيادك جامعة هارفارد على حياتك؟ لنجيب على هذا السؤال لا بد أن نعرف ماذا سيحدث لك عندما تلتحق بهارفارد – وماذا يحدث لك عندما لا تلتحق بهارفارد. من الواضح أننا لا نستطيع أن نمتلك بيانات عن هاتين الحالتين. لكن مع هذا يجد الباحثون العباقرة وسائل لمقارنة معالجة ما (كارتيادِ هارفارد على سبيل المثال) مع نقيضه، وهو ما سيحدث في غياب تلك المعالجة.

لشرح تلك النقطة، لنتأملْ هذا السؤال الذي يبدو بسيطًا: هل تَرْدَعُ زيادةُ أعداد ضباط الشرطة في الشوارع الجريمةَ؟ ذلك سؤال مهم من الناحية الاجتماعية؛ لأن الجريمة تفرض على المجتمع نفقات باهظة. لو كان تعاظم الحضور الشرطي يقلل من الجريمة، سواء عن طريق الردع أو بالقبض على الأشرار وسجنهم، فإن الاستثمار في زيادة أفراد الشرطة سيأتي بعائدات عظيمة. لكن من ناحية أخرى فإن رجال الشرطة باهظو التكاليف نسبيًّا، ولو كان تأثيرهم على الجريمة ضعيفًا أو لو لم يكن لهم أي تأثير على الإطلاق، فلربما كان من الأفضل أن يستغل المجتمع موارده في مكان آخر (ربما كان الأفضل هو الاستثمار في تكنولوجيا مكافحة الجريمة ككاميرات المراقبة).

يكمن التحدي هنا في كون سؤالنا هذا الذي يبدو بسيطًا – ما التأثير السببي لزيادة عناصر الشرطة على الجريمة؟ – قد اتضح أنه تصعب الإجابة عليه جدًّا. وبوصولنا لهذه

النقطة المتقدمة في الكتاب، لا بد أنك تعرف الآن أننا لا نستطيع أن نجيب عن هذا السؤال فقط بالتقصي حول إذا ما كانت الدوائر القضائية، التي يزيد فيها عدد رجال الشرطة في مقابل كل مواطن، تنخفض فيها معدلات الجريمة. فزيوريخ تختلف عن لوس آنجيليس. بل وحتى المقارنة بين المدن الأمريكية الكبرى ستكون معيبة بشدة، فَلوس آنجيليس ونيويورك وهيوستن وميامي وديترويت وشيكاجو، كلها أماكن مختلفة ذات ديمغرافيات وتحديات إجرامية مختلفة.

سيقتضي منهجنا المعتاد أن نحاول تحديد معادلة الانحدار التي تتحكم في كل هذه الاختلافات. للأسف، حتى تحليل الانحدار المتعدد لن ينقذنا هنا. لو حاولنا تفسير معدلات الجريمة (متغيرنا الاعتمادي) عن طريق استخدام عدد رجال الشرطة في مقابل كل مواطن كمتغير مفسر (بالإضافة إلى ضوابطنا الأخرى)، فسنجد نفسنا أمام مشكلة سببية عكسية حادة. إننا نمتلك سببًا نظريًّا قويًّا للاعتقاد بأن زيادة أعداد رجال الشرطة في الشوارع من شأنه تقليل الجرائم، لكنه من الممكن جدًّا كذلك أن تكون زيادة الجرائم هي التي «سببت» زيادة رجال الشرطة، حيث إن المدن التي تشهد موجات جرائم سوف تقوم بتعيين المزيد من رجال الشرطة. قد نجد بكل سهولة رابطًا إيجابيًّا، وإن كان مضللًا، بين الجريمة ورجال الشرطة: فالأماكن ذات التواجد الشرطي الأكثر كثافة هي التي تعاني أسوأ المشكلات الإجرامية. وبالطبع فإن الأماكن التي يتواجد فيه الأطباء بكثافة هي التي يتركز فيها وجود المرضى. هؤلاء الأطباء لا يصيبون الناس بالمرض، وإنما وضعوا حيث الناس في أشد الاحتياج لهم (وفي نفس الوقت ينتقل المرضى إلى الأماكن التي تتيح لهم الحصول على العناية الطبية المناسبة). إنني لأشك في أنه في فلوريدا أعداد كبيرة من إخصائيي الأورام وإخصائيي القلب، وأنهم لو أبعدوا عن الولاية لما جعل ذلك المسنين من المواطنين أكثر صحة.

أهلًا بك في عملية تقييم البرنامج، وهي عملية ننشد من خلالها قياس التأثير السببي لتداخلٍ ما – أشياء من قبيل عقَّار جديد لعلاج السرطان أو حتى برنامج لتوظيف المتسربين من المدارس أو زيادة أعداد رجال الشرطة في الشوارع. ذلك التداخل الذي نُعنى به يسمى دائمًا «معالجة»، رغم أن هذه الكلمة تستخدم بشكل متسع النطاق ضمن السياق الإحصائي أكثر منها في اللغة المعتادة. فالمعالجة قد تعني المعالجة بالمعنى الحرفي للكلمة، كما هو الحال في بعض التداخلات الطبية، أو قد تكون أمرًا على غرار الالتحاق بالجامعة، أو أن تتلقى تدريبًا مهنيًّا عقب إطلاق سراحك من السجن. المقصد هنا أن نقوم بعزل تأثير عامل مفرد، وفي الظروف المثالية نرمي لأن نعرف كيف هو حال مجموعة تتلقى معالجة مقارنة بمجموعة أخرى تطابقها في كل الجوانب عدا تلك المعالجة.

تقييم البرنامج يقدم لنا مجموعة من الأدوات لعزل تأثير المعالجة عندما يكون السبب

والتأثير مراوغين. فيما يلي نسرد كيف قام جوناثان كليك الباحث بجامعة بنسلفانيا، وآلكزاندر تاباروك الباحث بجامعة جورج ميسون، بدراسة كيف تؤثر زيادة أعداد رجال الشرطة في الشوارع على معدلات الجريمة. لقد استفادت استراتيجيتهم البحثية من نظام الإنذار بالإرهاب. وعلى وجه التحديد نقول إن واشنطن العاصمة تستجيب لأيام «التهديدات الخطيرة» بوقوع حوادث إرهابية بأن تنشر المزيد من رجال الشرطة في مواقع معينة من المدينة، حيث إن العاصمة بطبعها هدف للإرهابيين. يمكننا أن نفترض عدم وجود علاقة بين جرائم الشوارع وتهديدات الإرهاب، لذا فإن تلك الزيادة في التواجد الشرطي لا علاقة لها بمعدل الجرائم المتعارف عليه، أو هي زيادة «خارجية». كانت أعلى الرؤى البحثية قيمة هي ملاحظة تلك التجربة الطبيعية التالية: ماذا يحدث للجريمة العادية في تلك الأيام التي «ترتفع» فيها التهديدات الإرهابية؟

الإجابة: كان عدد الجرائم التي ارتكبت عندما كان الإنذار بالتهديدات الإرهابية برتقالي اللون (إنذارًا عاليًا ومزيدٌ من رجال الشرطة على الأرض) أدنى بنحو ٧ بالمائة، منها وقت أن كان الإنذار بالتهديدات الإرهابية أصفر اللون (وهو ما يعني إنذارًا مرتفعًا لكن دون أي احتياطات زائدة من قوات إنفاذ القانون). كما وجد واضعو الدراسة أن معدل الجريمة يكون في أدنى درجاته في الدوائر الشرطية التي تكون في أعلى درجات استنفارها في أيام التهديدات الإرهابية (لأنها تشمل البيت الأبيض ومقر السلطة التشريعية والمتنزه الوطني). أهم الفوائد التي نخرج بها مما سبق هي أننا صرنا قادرين على الإجابة عن أسئلة خادعة، وإن كانت ذات دلالات اجتماعية – وكل ما علينا أن نجيد تلك الإجابات. فيما يلي بعض أكثر مناهج عزل تأثير المعالجة شيوعًا.

التجارب الضابطة المجراة عشوائيًّا. أكثر طرق خلق مجموعات الضبط والمعالجة مباشرة هي – تأهب لها – بخلق مجموعات ضبط ومعالجة. في هذا المنهج تحدِّيان كبيران. أولهما وجود العديد من أنواع التجارب التي لا نستطيع إجراءها على البشر. وهذا القيد (كما آمل) لن ننفك منه في أي وقت قريب. ونتيجة لهذا فإننا لا نستطيع أن نجري تجارب ضابطة على البشر إلا إذا تبدى أمامنا سبب يدفعنا للاعتقاد بأن تأثير تلك المعالجة يُنتظر منه نتائج إيجابية. لكن لا ينطبق هذا على غالبية الحالات (كما في «معالجات» على غرار اختبار عقاقير جديدة أو التسرب من التعليم الثانوي)، ولذلك فنحن بحاجة إلى تلك الاستراتيجيات التي نعرض لها خلال هذا الفصل.

ثانيًا، هناك من التغاير بين البشر ما يزيد عنه لدى فئران التجارب. وتأثير المعالجة الذي نختبره قد يسهُل التشويش عليه من قِبل تغايرات أخرى في المعالجة ومجموعات الضبط؛ فإنك ملزم بأن تقبل طوال الناس وقصارهم، ومرضاهم وأصحاءهم، وذكورهم وإناثهم، والمجرمين منهم، ومدمني الكحوليات، والمستثمرين المصرفيين منهم، وهلم جرًّا.

كيف لنا أن نتأكد من أن الاختلافات في هذه الخصائص لن تفسد نتائجنا؟ لديَّ لك أخبار جيدة، وهي أن هذا واحد من الأمثلة النادرة في الحياة، والتي يتطلب فيها أفضل المناهج أقل قدر من العمل. فالطريقة المثلى لخلق أي مجموعة معالجة ومجموعة ضبط هي بتوزيع المشاركين في الدراسة على المجموعتين بشكل عشوائي. يكمن جمال التوزيع العشوائي في كونه سوف يوزع المتغيرات غير ذات الصلة بالمعالجة بتساوٍ تقريبي على المجموعتين – سواء الخصائص الواضحة كالجنس والعرق والعمر ومستوى التعليم، أو الخصائص غير الملحوظة التي لو لم توزع بتساوٍ لأفسدت النتائج.

فكر في الأمر، لو كان في عينتنا المرتقبة ١٠٠٠ أنثى، ثم قسمنا العينة عشوائيًّا إلى مجموعتين، فإن الناتج المرجح لهذا هو أن كل مجموعة منهما سوف تحتوي على ٥٠٠ أنثى. بالطبع لا نستطيع أن نتوقع تلك القسمة بالضبط، لكن مرة أخرى نجد الاحتمالات صديقتنا. فاحتمالات أن يتوزع قدر غير متناسب على إحدى المجموعتين ضعيفة (أو أي قدر غير متناسب من الأفراد الذين يتسمون بأي خاصية أخرى). فمثلًا تأمل المجموعة التي ذكرناها لتوًّا والتي تحتوي على ١٠٠٠ امرأة. عندما نقسم تلك المجموعة عشوائيًّا إلى عينتين، فهناك احتمال قدره ١ بالمائة أن تحتوي إحداهما على أقل من ٤٥٠ امرأة. ومن الواضح أنه كلما كبرت العينات صار التوزيع العشوائي أكثر فعالية في خلق مجموعتين متشابهتين على نطاق واسع.

عادة ما تطمح الاختبارات الطبية لإجراء تجارب ضابطة ذات انتقاء عشوائي. عادة ما تكون تلك الاختبارات الطبية مزدوجة التعمية، بمعنى أنه لا المريض ولا الطبيب يعلم من يتلقى المعالجة ومن يتلقى العقَّار الوهمي. وهو الأمر الذي يستحيل بالطبع مع معالجات معينة كالإجراءات الجراحية (فجرَّاح القلب، كما آمل، سوف يعرف أيُّ المرضى سيخضع لجراحة فتح مجرًى جانبي للشريان التاجي). لكن حتى مع الإجراءات الجراحية يمكن أن نمنع المرضى من معرفة إذا ما كانوا في مجموعة المعالجة أو مجموعة الضبط. واحدة من الدراسات المفضلة لديَّ تتعلق بتقييم نوع من أنواع جراحات الركبة التي تهدف إلى تقليل الألم. خضعت مجموعة المعالجة لتلك الجراحة. أما مجموعة الضبط فقد خضعت لجراحة «وهمية»، وفيها قام الجراح بعمل ثلاثة جراح قطعية، «وتظاهر بأنه يجري» الجراحة*. وقد اتضح أن الجراحة الحقيقية ليست أكثر فعالية من الجراحة الوهمية من حيث تخفيف آلام الركبة[١]

يمكن أن تستخدم التجارب التي تتسم بالتوزيع العشوائي في اختبار ظواهر مثيرة للاهتمام. فمثلًا هل تحسن الصلوات التي يصليها الآخرون للمرضى في تحسين حالتهم الصحية بعد إجرائهم لعمليات جراحية؟ العقلانيون من الناس يرون الدين برؤًى شديدة

* كان المشاركون يعلمون أنهم مشاركون في تجربة سريرية، وأنهم قد يخضعوا لجراحة وهمية.

التباين، لكن إحدى الدراسات التي نشرت في دورية American Heart Journal قامت بإجراء دراسة ضابطة بحثت فيما إذا كان المرضى المتعافون من جراحة فتح المجرى الجانبي للشريان التاجي سيتعرضون لمضاعفات أقل في مرحلة ما بعد الجراحة إذا ما صلَّت مجموعة كبيرة من الغرباء لأجل صحتهم وتعافيهم السريع[٢]. شارك في هذه الدراسة ١٨٠٠ مريض وأعضاء لثلاث مؤسسات دينية من مختلف أنحاء البلاد. تم تقسيم أولئك المرضى، الذين خضعوا جميعًا لجراحة فتح مجرًى جانبي للشريان التاجي، إلى ثلاث مجموعات: واحدة منها لم يصلِّ لها أحد، والثانية صُلي لأجلهم وأُخبروا بذلك، أما الثالثة فقد صُلي لأجلها لكن أفرادها أُخبروا أنهم قد تتم الصلاة لأجلهم وقد لا تتم (وهو الأمر الذي يحكم تأثير الصلاة الوهمية). وفي الوقت نفسه طلب من أعضاء المؤسسات الدينية أن يصلوا لأجل مرضى معينين باسمهم الأول وأول حرف من لقبهم (تشارلي دبليو. مثلًا). أعطي المصلون الحرية فيما يتعلق بكيفية الصلاة شرط أن تتضمن عبارة «لأجل جراحة ناجحة يتبعها شفاء سريع صحي وبلا مضاعفات».

ثم ماذا؟ هل ستكون الصلاة هي الحل الفعال من ناحية التكلفة للتحديات التي تواجه الصحة العامة في أمريكا؟ لا على الأرجح. لم يجد البحث أي اختلاف بين معدلي المضاعفات خلال ثلاثين يومًا من الجراحة لأولئك الذين صُلي لأجلهم وأولئك الذين لم يُصَلَّ لهم. أشار ناقدو هذه الدراسة إلى متغير محذوف محتمل، ألا وهو الصلوات الآتية من مصادر أخرى. لخصت النيويورك تايمز الأمر بقولها: «قال الخبراء إن الدراسة لم تستطع التغلب على ما قد تكون أكبر عَقَبة تواجه دراسة عن الصلاة، ألا وهي: عدم معرفة قدر الصلوات التي يتلقاها كل شخص من الأصدقاء والعائلات ورجال الدين حول العالم، والذين يصلون لكل من هو مريض أو يحتضر».

إن التجريب على البشر قد يزج بك في السجن، أو ربما يسوقك لأن تقف أمام هيئة محكمة جرائم دولية. عليك أن تعي لذلك. لكن رغم ذلك لا يزال هناك متسع في العلوم الاجتماعية لإجراء تجارب ضابطة ذات انتقاء عشوائي على «بشر». واحدة من أشهر التجارب وأكثرها تأثيرًا كانت تجربة مشروع تينيسي STAR، والتي اختبرت تأثير صغر حجم الصفوف الدراسية على تعلم التلاميذ. إن العلاقة بين حجم الصف الدراسي والتعلم هي علاقة مهمة جدًّا. والأمم في جميع أرجاء العالم تكافح لأجل تحسين مخرجاتها التعليمية. لو كان صغر حجم الصفوف الدراسية كفيلًا بجعل التعليم أكثر فعالية، مع تثبيت بقية العوامل، فإنه يتعين على المجتمع أن يقوم بتوظيف المزيد من المعلمين لتقليص أحجام الصفوف. لكننا نجد في الوقت نفسه أن توظيف المعلمين أمر مكلف، فلو كان التلاميذ الذين يدرسون في صفوف صغيرة الحجم يُبلون بلاءً أفضل لأسباب لا تتعلق بحجم الصفوف، فسنكون حينها نهدر كمًّا ضخمًا من الأموال.

من المدهش أن العلاقة بين حجم الصف وإنجازات التلاميذ هي علاقة يصعُب دراستُها. فالمدارس التي تتميز بصغر حجم صفوفها هي في الغالب مدارس ذات موارد ضخمة، بمعنى أن تلاميذها ومعلميها على السواء يرجَّح أنهم يختلفون عن نُظرائهم في المدارس ذات الصفوف الأكبر حجمًا. وفي داخل كل مدرسة قد تكون تلك الصفوف أصغر حجمًا لسبب معين. فقد يقوم ناظر إحدى المدارس بتوزيع التلاميذ صعاب المراس على الصفوف الأقل حجمًا، وفي هذه الحالة قد نجد ارتباطًا سلبيًّا زائفًا بين الصفوف الأصغر حجمًا وإنجاز التلاميذ. أو قد يختار المعلمون المخضرمون أن يدرِّسوا للصفوف الأصغر حجمًا، وفي هذه الحالة قد تنبع الفائدة من الصفوف الأصغر حجمًا من المعلمين الذين يختارون أن يدرِّسوهم، لا من صغر نسبة التلاميذ الذين يدرِّسهم هؤلاء المعلمون.

بداية من عام ١٩٨٥ أجرى مشروع تينيسي STAR تجربة ضابطة لاختبار تأثيرات الصفوف الأصغر حجمًا[٢] (كان لامار آلكزاندر وقتها حاكم تينيسي، وصار بعدها وزيرًا للتعليم في إدارة الرئيس جورج بوش الأب). تم تقسيم تلاميذ مرحلة الحضانة في تسع وسبعين مدرسة مختلفة بشكل عشوائي إلى صفوف إما صغيرة (سعتها من ١٣ إلى ١٧ تلميذًا) أو عادية (سعتها من ٢٢ إلى ٢٥ تلميذًا)، أو صفوف عادية بكل منها معلم ومعلم مساعد. وقد وُزع المعلمون كذلك عشوائيًّا على صفوف مختلفة. بقي التلاميذ في نفس نوع الصف الذي وُزعوا إليه حتى الصف الثالث. لكن اختلاف الواقع الحياتي لأولئك التلاميذ قد أخلَّ بذلك الانتقاء العشوائي؛ فقد أُدخل بعض التلاميذ في هذه المنظومة في منتصف التجربة، وغادرها آخرون. ونُقل بعض التلاميذ من صف لآخر لأسباب تأديبية، وأقدم بعض الآباء على الضغط كي يُنقل أبناؤهم إلى صفوف أصغر حجمًا، وهلم جرًّا.

لكن يظل مشروع STAR الاختبار الوحيد ذا الانتقاء العشوائي لتأثير الصفوف الدراسية الأصغر حجمًا. وقد اتضح أن نتائجه لها دلالات إحصائية واجتماعية. فبشكل عام كان أداء تلاميذ الصفوف الصغرى أفضل بقدر ١٥ انحرافًا معياريًّا في الاختبارات الموحدة من مثيله لدى تلاميذ الصفوف المعتادة، وجاء أداء التلاميذ سود البشرة في تلك الصفوف الصغرى أفضل بالضعف. والآن إلى الأنباء السيئة. كُلفت تجربة مشروع STAR نحو ١٢ مليون دولار. وكُلفت دراسة تأثير الصلوات على مضاعفات ما بعد الجراحة ٢٫٤ مليون دولار. فأفخر الدراسات كأفخر كل شيء، باهظة التكاليف.

التجربة الطبيعية. لسنا كلنا نملك الملايين من الدولارات التي لا حاجة لنا بها كي ننفقها على خلق تجربة كبيرة ذات توزيع عشوائي. لكن البديل الأكثر اقتصادًا يتمثل في استغلال تجربة طبيعية، وهو الأمر الذي يحدث عندما ينتُج عن بعض الظروف الطبيعية شيء يقارب شكل التجربة الضابطة ذات التوزيع العشوائي. وهو الأمر الذي انطبق على مثال شرطة واشنطن العاصمة الذي ذكرناه أول الفصل. فأحيانًا ما تخلق الحياة مجموعة

معالجة ومجموعة ضابطة عن طريق الصدفة، وعندما يحدث ذلك نرى الباحثين متلهفين للقفز على النتائج. لنتأمل الرابط المدهش، مع كونه معقدًا، والذي يربط بين مستوى التعليم وطول العمر. فمن يتلقون مستوًى أعلى من التعليم يغلب أن يعيشوا لعمر أطول، حتى مع تثبيت عوامل كمستوى الدخل ومدى الرعاية الصحية التي يحصلون عليها. فقد أشارت النيويورك تايمز إلى أن «العامل الاجتماعي الوحيد، الذي يتفق الباحثون على كونه يرتبط ارتباطًا مستمرًّا بالعيش لعمر أطول في كل دولة درس فيها، هو مستوى التعليم. فهو أقوى من عامل العِرق، وهو يطمس أي تأثير لعامل مستوى الدخل»[٤]. لكن حتى الآن ليس هذا سوى علاقة ارتباطية. هل يؤدي مزيد من التعليم، مع تثبيت العوامل الأخرى، إلى صحة أفضل؟ لو فكرت في التعليم نفسه على أنه «معالجة»، فهل من المحتمل إذا ما تلقيت المزيد من التعليم يجعلك ذلك تعيش أطول؟

قد يبدو ذلك سؤالًا تكاد تستحيل دراسته، حيث إن مَن يختارون أن يتلقوا المزيد من التعليم يختلفون عمن لا يختارون ذلك. فالفارق بين خريجي المدارس الثانوية وخريجي الجامعات لا يقتصر على كونه مجرد أربع سنوات دراسية. فمن الممكن جدًّا أن تكون هناك بعض الخصائص غير الملحوظة، والتي يشترك فيها من يواصلون التعلم، وهي التي يمكن أيضًا أن تفسر ارتفاع معدلات أعمارهم. لو صحَّت تلك الحالة فلن يحسُن تقديم المزيد من التعليم لمن اختاروا تعليمًا أقل من حالتهم الصحية. فلن يكون تحسن الصحة نتاجًا للتعليم التدريجي، وإنما سيكون نتاجًا لنوعية هؤلاء الناس الذين ينشدون التعليم التدريجي.

لا يمكننا أن نجري تجربة ذات توزيع عشوائي لأجل حل هذه المعضلة، لأن ذلك سيستلزم جعل بعض المشاركين يتركون المدرسة في وقت أبكر مما يريدون. (حاوِل أن تشرح لشخص ما أنه لا يستطيع أن يلتحق بالجامعة – أبدًا – لأنه عضو في مجموعة ضابطة)؛ لذا فإن الاختبار الوحيد الممكن للتأثير السببي للتعليم على طول العمر، لا بد أن يكون نوعًا من أنواع التجارب التي أُجبر فيها قطاع كبير من المجتمع لأن يظل في المدرسة لوقت أطول مما قد يكون أعضاؤه قد اختاروا. تلك تجربة نجدها مقبولة أخلاقيًّا على الأقل، وذلك لأننا ننتظر منها تأثيرًا معالجًا إيجابيًّا. لكننا مع هذا لا نستطيع أن نجبر الأطفال على أن يستمروا في المدارس، ليست تلك بالطريقة الأمريكية.

بل إنها الطريقة الأمريكية، ففي كل ولاية نجد ضربًا من ضروب القوانين التي تُلزم المواطنين بحد أدنى من التعليم، وتلك القوانين قد تغيرت في مراحل تاريخية مختلفة. وهذا النوع من أنواع التغير الخارجي في التحصيل التعليمي – بمعنى أنه لم يسببه الأفراد الخاضعون للدراسة – هو بالضبط الأمر الذي يجعل الباحثين يمتلئون حماسًا. كانت أدريانا ليراس موني، طالبة الدراسات العليا في جامعة كولومبيا، هي التي رأت في

الأمر موضوعًا بحثيًّا محتملًا من حيث إن ولايات أمريكية مختلفة قد غيرت قانون الحد الإلزامي للتعليم في مراحل تاريخية مختلفة. فقد رجعت إلى التاريخ ودرَست العلاقة بين الأزمنة التي غَيَّرت فيها تلك الولاياتُ قوانينَ الحد الإلزامي للتعليم، وما تلا ذلك من تغيرات في متوسط الأعمار بنفس تلك الولايات (من خلال التنقيب داخل كمٍّ هائل من بيانات إحصائيات تعداد السكان). لكنها مع ذلك وجدت نفسها أمام تحدٍّ منهجي، فإذا كان سكان ولاية ما قد ازداد متوسط أعمارهم بعدما رفعت الولاية من حدها الإلزامي للتعليم، لا نستطيع أن نعزو هذه الزيادة لزيادة مستوى التعليم. إن متوسط العمر يرتفع بشكل عام بمرور السنين. فقد ازداد متوسط العمر عام ١٩٠٠ عنه في عام ١٨٥٠، بصرف النظر عما فعلت الولايات.

لكن رغم هذا وجدت ليراس موني مجموعة ضابطة طبيعية، ألا وهي تلك الولايات التي لم تغيِّر قوانين الحد الإلزامي للتعليم. وقد صار عملها أقرب لتجربة معملية عملاقة، أُجبر فيها سكان إيلينوي على أن يظلوا في المدارس لسبعة أعوام، بينما تَرك جيرانُهم في إنديانا المدرسةَ بعد ستة أعوام. الاختلاف هنا هو أن هذه التجربة الضابطة قد تأتَّت من صدفة تاريخية – ومن هنا جاء مصطلح «التجربة الطبيعية».

ما الذي حدث؟ لقد ازداد متوسط العمر المتوقع للبالغين الذين وصلوا لسن الخامسة والثلاثين بمعدل عام ونصف باستمرارهم في التعليم لسنة إضافية[٥] استُنسخت نتائج ليراس موني في دول أخرى، خَلقت فيها تنوعاتُ قوانين الحد الإلزامي للتعليم تجربةً طبيعية مشابهة. لكن ما زال في الأمر مجال للتشكك، فإننا لا نفهم حتى الآن الآلية التي تؤدي بها سنةُ تعليمٍ إضافيةٌ إلى حياة أطول.

التحكم غير المكافئ. أحيانًا يكون أفضل خيار متاح لدراسة تأثير المعالجة هو خلق مجموعتَي معالجة وضبط غير موزعتين عشوائيًّا. هنا نأمل أو نتوقع أن تكون المجموعتان متشابهتين على نطاق واسع، حتى ولو لم تمنحنا الظروفُ الرفاهيةَ الإحصائية للقيام بالتوزيع العشوائي. لكن الخبر الجيد هنا أننا نمتلك مجموعة معالجة ومجموعة ضبط. أما الخبر السيء فهو أن أي توزيع غير عشوائي من المحتمل أن يخلق انحيازًا. فقد تكون هناك اختلافاتٌ غير ملحوظة بين مجموعتَي الضبط والمعالجة ذاتُ صلة بتوزيع المشاركين على مجموعة دون الأخرى. ومن هنا جاءت تسمية «التحكم غير المكافئ».

لكن حتى مجموعة الضبط غير المكافئ تظل أداة ذات نفع كبير أيضًا. دعنا نتفكر في ذلك السؤال الذي يطرحه عنوان هذا الفصل: هل هناك ميزة حياتية مهمة للالتحاق بكلية أو جامعة تنتقي الملتحقين بها؟ من الواضح أن خريجي هارفارد وبرنستون ودارتموث من جميع أنحاء العالم يُبلون بلاءً حسنًا. فهُم في متوسطهم يحوزون أموالًا أكثر ويحظون بفرص حياتية أكثر من الطلبة الذين تخرجوا في معاهد تعليمية أقل انتقاءً لملتحقيها.

(وَجَدَت دراسةٌ أجراها موقع PayScale.com عام ٢٠٠٨ أن وسيط ما يتقاضاه خريجو دارتموث ذوو الخبرة العملية التي تتراوح ما بين عشر سنين وعشرين سنة يبلغ ١٣٤٠٠٠ دولار أمريكي، وهو الأعلى بين خريجي أي مؤسسة تعليمية من غير الحاصلين على شهادات دراسات عليا، وتَلَتْها برنستون في المركز الثاني بوسيط قدره ١٣١٠٠٠ دولار أمريكي)[٦] لعلك الآن، كما آمل، تعرف أن هذه المبالغ الباهظة لا تشي بأي شيء عن قيمة التعليم في دارتموث وبرنستون. فهؤلاء الطلبة الذين يتقدمون للالتحاق بدارتموث وبرنستون هم طلبة موهوبون منذ تقدُّمهم لها، ولذلك قُبلوا فيهما. وهم على الأرجح سوف يُبلون بلاءً حسنًا في الحياة بصرف النظر عن الكلية التي تخرجوا فيها.

ما نجهله هنا هو تأثير المعالجة الذي يترتب على ارتياد جامعات على غرار هارفارد وَييل. هل ينجح خريجو مؤسسات الصفوة تلك في الحياة لأنهم كانوا ذوي موهبة استثنائية وقت أن خَطَوا إلى داخل الحرم الجامعي؟ أم إن تلك الكليات والجامعات تكون ذات قيمة مضافة عندما تأخذ هؤلاء الأفراد الموهوبين وتجعلهم أكثر إنتاجية؟ أم الاثنان معًا؟

لا نستطيع أن نجري تجربة ذات توزيع عشوائي تجيب عن هذا السؤال. فالقليل من تلاميذ المدارس الثانوية سيوافقون على أن يوزَّعوا عشوائيًّا على إحدى الكليات، ولا هارفارد أو دارتموث ستتحمسان لأن تقبلا طلبة يوزعون عليهما بشكل عشوائي. قد يبدو في الظاهر أننا لا نملك أية آلية لاختبار قيمة تأثير المعالجة. الذكاء هو الذي سينقذنا هنا، فقد وجد الباحثان الاقتصاديان ستايسي دايل وآلان كروجر طريقةً للإجابة عن هذا السؤال من خلال استغلال* حقيقة أن الكثير من الطلبة يتقدمون بطلبات للالتحاق بأكثر من جامعة[٧]. وبعض هؤلاء الطلبة يُقبلون في جامعات ذات شأن عالٍ ويَختارون أن يلتحقوا بنفس تلك الجامعات، بينما يُقبل آخرون في جامعات ذات شأن عالٍ لكنهم يَختارون أن يلتحقوا بجامعات أو كليات أقل شأنًا. وجدناها! صارت لدينا الآن مجموعة معالجة (أولئك الذين التحقوا بالكليات والجامعات الأعلى شأنًا) ومجموعة ضبط غير مكافئ (أولئك الطلبة الذين كانوا بالقدر الكافي من الموهبة لأن يُقبلوا في تلك الجامعات الأعلى شأنًا لكنهم فضَّلوا أن يلتحقوا بجامعات أقل منها شأنًا)**.

---

* يحب الباحثون أن يستخدموا كلمة «استغلال exploit»، فهي ذات معنًى محدد فيما يتعلق بانتهاز فرصة ذات صلة ببعض البيانات. فمثلًا عندما يعثر الباحثون على تجربةٍ طبيعيةٍ ما، من التي تخلق مجموعة ضابطة ومجموعة معالجة، حينها سوف يصفون كيف يخططون لـ «استغلال التغاير الحادث في البيانات».

** هناك احتمال لحدوث انحياز هنا. فتلاميذ كلتا المجموعتين ذوو موهبة كافية لأن يُقبلوا في الجامعات الأعلى شأنًا. لكن مجموعة من التلاميذ فضلوا الالتحاق بجامعةٍ ما والمجموعة الأخرى لم تفضل ذلك. تلاميذ تلك المجموعة الذين فضلوا الالتحاق بالجامعات الأقل شأنًا قد يكونون أقل تحفيزًا وأقل اجتهادًا ومختلفين بطرائق أخرى قد لا يمكننا أن نلاحظها. حتى لو وجد دايل وكروجر أن الطلبة الذين التحقوا بتلك الجامعات العالية الشأن قد غنموا من حياتهم أكثر مما غنم أولئك الذين قُبلوا فيها لكنهم التحقوا بغيرها، لن يكون بإمكاننا أن نتيقن من كون هذا الاختلاف راجعًا إلى تلك الكلية الرفيعة أم إلى نوع ذلك الطالب الذي أقدم على =

درس دايل وكروجر البيانات الطولية التي تخص ما يتقاضاه أفراد المجموعتين. ليست تلك بمقارنة مثالية من نوع مقارنة تفاحات بتفاحات، ومن الواضح أن الأموال التي يتقاضونها ليست بالمخرج الحياتي الوحيد المهم، لكن النتائج التي خلصوا إليها من شأنها أن تريح بال طلاب المدارس الثانوية وذويهم. فالطلاب الذين التحقوا بالجامعات الأعلى شأنًا يتقاضون نفس ما يتقاضاه الطلاب الذين يماثلونهم في القدرات والذين التحقوا بجامعات أقل شأنًا تقريبًا. وتَمثَّل الاستثناء الوحيد في حالة الطلاب المنتمين إلى أسر محدودة الدخل، والذين زاد ما تقاضوه بعدما التحقوا بجامعات أو كليات أعلى شأنًا. إن المنهج الذي اتبعه دايل وكروجر ما هو إلا طريقة أنيقة للفصل بين تأثير المعالجة (والذي يتمثل في قضاء أربع سنوات في مؤسسة تعليمية استثنائية)، وتأثير الانتقاء (قبول أكثر الطلاب موهبة في هذه المؤسسات). وفي تلخيص لهذا البحث كتبه آلان كروجر في صحيفة النيويورك تايمز، أجاب بشكل غير مباشر عن السؤال الذي عَنْوَنَّا به هذا الفصل «عليك أن تدرك أن دافعيتك وطموحك ومواهبك هي التي تحدد مدى نجاحك أكثر من اسم الكلية التي ستلتحق بها أو الشهادة التي ستحصل عليها»[٨].

الاختلاف في الاختلافات. من أفضل الطرق لملاحظة السبب والتأثير هو أن يقوم المرء بفعل شيء، ثم ملاحظة ما سيحدث. فهذا هو ما يفعله الرضع والأطفال الصغار (والبالغون أحيانًا) ليتعلموا حقائق العالم. فقد تَعلَّم أطفالي بسرعة شديدة أنهم إذا قذفوا بلقيمات من الطعام في أي مكان بالمطبخ (السبب)، فسيهرع الكلب نحوها مسرعًا (التأثير). ومن المفترض أن نفس قوة الملاحظة تلك قادرة على إثراء بقية الحياة بالمعلومات. فلو قمنا بعمل استقطاعات ضريبية فتَحسَّن الاقتصاد بعدها، فلا بد أن هذه الاستقطاعات كانت تصرفًا مسئولًا.

ربما. لكن الفخ الوحيد المحتمل في هذا المنهج هو أن الحياة تميل لأن تكون أكثر تعقيدًا من مجرد الرمي بقطع من الدجاج في مكانٍ ما من المطبخ. نعم، لقد فرضنا ضرائب في وقت محدد، لكن في نفس ذلك الوقت كانت هناك «تداخلات» أخرى تتكشف خلال نفس تلك المرحلة الزمنية تقريبًا: ازداد عدد النساء اللاتي التحقن بالدراسة الجامعية، وصارت شبكة الإنترنت والتكنولوجيا الحديثة ترفعان من إنتاجية العمالة الأمريكية، وكانت العملة الصينية قد انخفضت قيمتها، وأقال فريق شيكاجو كابز للبيسبول مديرهم العام، وهلم جرًّا. فأيًّا كان ما حدث بعد فرض الضرائب لا يمكن أن يعزى فقط إلى فرض الضرائب.

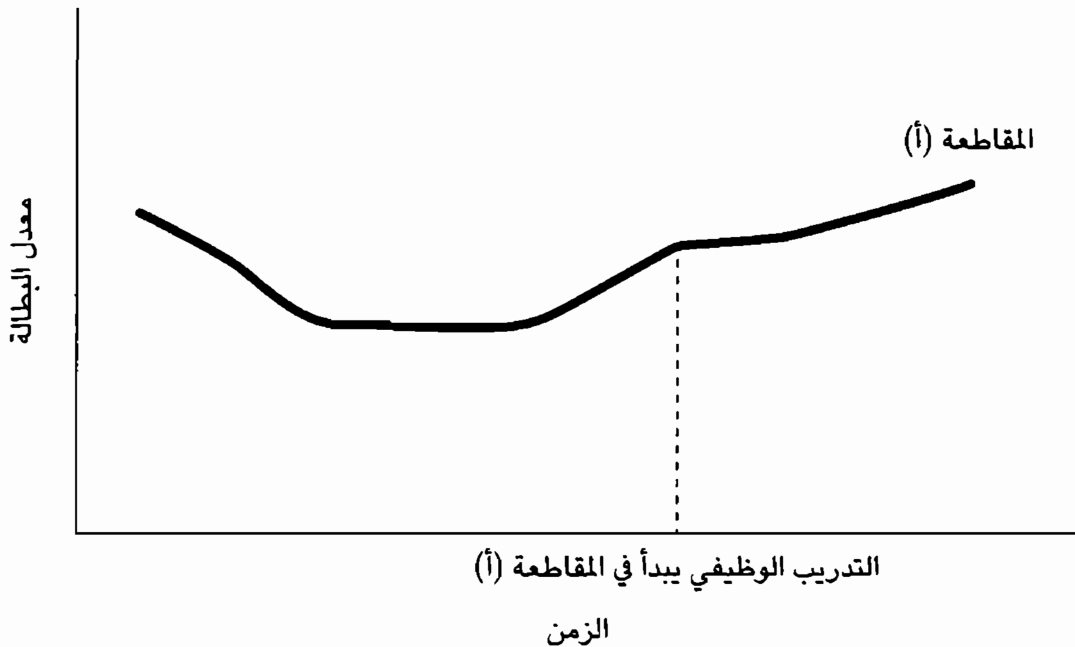
= الالتحاق بها عندما أتيح له ذلك الخيار. لكن اتضح أن هذا الانحياز المحتمل لم يكن مهمًّا في دراسة دايل وكروجر؛ وذلك بسبب توجهها. وجد دايل وكروجر أن الطلبة الذين ارتادوا تلك الجامعات الأعلى شأنًا لم تفُق مغانمهم مغانم رفاقهم الذين قبلوا فيها ولم يلتحقوا بها كثيرًا، رغم أن الطلبة الذين رفضوا الالتحاق بجامعة أعلى شأنًا قد تكون لديهم خصال تجعلهم يغنمون أقل من رفاقهم في الحياة بصرف النظر عن مستوى تعليمهم. هذا الانحياز يجعل النتائج تبالغ في الفوائد المالية لارتياد جامعة عالية الشأن جدًّا – وهو الأمر الذي اتضح كونه غير جوهري على الإطلاق.

فالتحدي الذي يكمن في تحليل من نوع «قبل وبعد» هو أنه ليس لمجرد كونِ حدثٍ ما سبق الآخر، فلا يعني هذا وجود علاقة سببية بين الاثنين.

يمكن أن يساعدنا منهج «اختلاف الاختلافات» في تحديد تأثيرات تداخُلٍ ما عن طريق القيام بأمرين. أولهما هو فحص بيانات «ما قبل» و«ما بعد» الخاصة بأي مجموعة كانت أو أي منطقة ما تلقت تلك المعالجة، مثل نسب البطالة في مقاطعة طَبَّقَت برنامجًا للتدريب الوظيفي. وثانيهما هو مقارنة هذه البيانات بنسب البطالة في نفس المرحلة الزمنية في مقاطعة مشابهة لم تطبِّق ذلك البرنامج.

الافتراض المهم هنا هو أن المجموعتان المستخدمتان في هذا التحليل قابلتان للمقارنة بشكل واسع عدا في المعالجة، ونتيجة لهذا يصير أي اختلاف دالٍّ بين المجموعتين قابلًا لأن يعزى إلى البرنامج أو السياسة التي نقيِّمها. فمثلًا لنفترض أن إحدى مقاطعات إيلينوي تطبِّق برنامجًا للتدريب الوظيفي طلبًا للقضاء على ارتفاع معدل البطالة. لكن في السنين التي تلت ذلك التطبيق، واصل معدل البطالة ارتفاعه، هل ذلك يجعل البرنامجَ برنامجًا فاشلًا؟ مَن علَّه يعرف؟

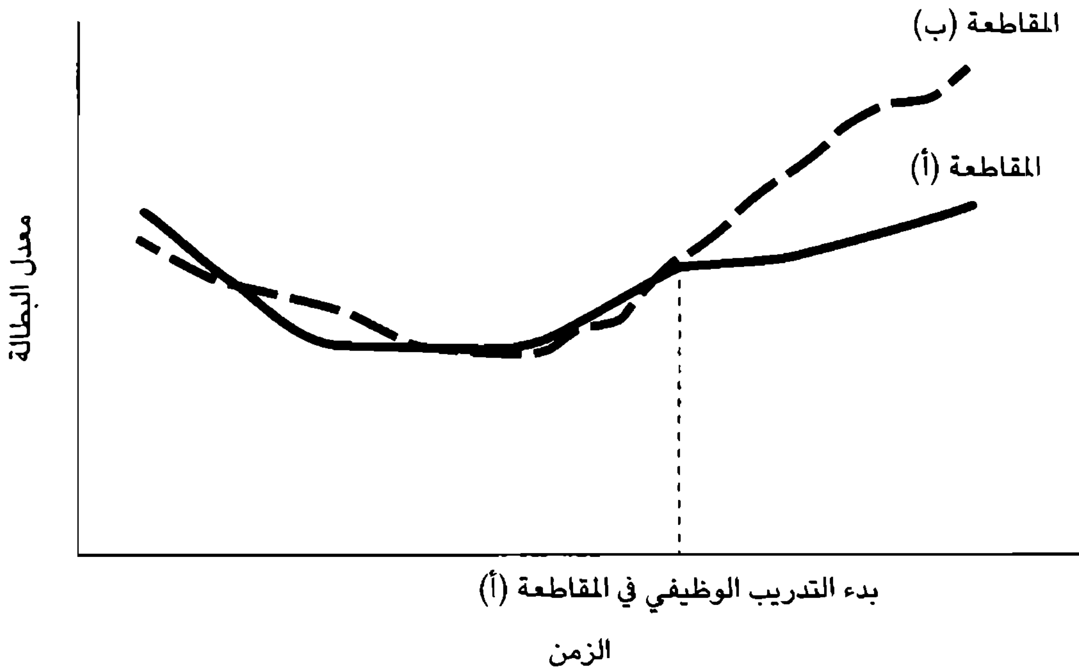
تأثير التدريب الوظيفي على معدل البطالة في المقاطعة (أ)



وربما كانت هناك من القوى الاقتصادية الفاعلة ما هي ذات تأثير جارٍ، ومن بينها الانكماش الاقتصادي الطويل المدى. إن منهج الاختلاف في الاختلافات يقوم بالمقارنة بين التغير الحادث في معدل البطالة بمرور الزمن في تلك المقاطعة التي نحن بصدد تقييمها، ومعدل البطالة في المقاطعة الجارة لها التي لا تطبِّق برنامج التدريب الوظيفي، ويجب

أن تكون المقاطعتان متشابهتين في جميع الوجوه الأخرى المهمة مثل مجالات الأعمال، والتركيبة الديمغرافية، وهلم جرًّا. كيف يتغير معدل البطالة في المقاطعة التي تطبِّق برنامج التدريب الوظيفي بمرور الزمن، نسبةً إلى المقاطعة التي لم تطبِّق ذلك البرنامج؟ يمكننا أن نستنتج، استنتاجًا منطقيًّا، تأثيرَ معالجة البرنامج من خلال المقارنة بين التغيرات الحادثة في المقاطعتين بطول المدى الزمني في الدراسة – «الاختلاف في الاختلافات». تمثِّل المقاطعة الأخرى في هذه الدراسة المجموعة الضابطة بشكل فعال، وهو الأمر الذي يسمح لنا أن نستفيد من البيانات التي جمعناها قبل التداخل وبعده. ولو حسنت المجموعة الضابطة، لتعرضت لنفس القوى الشديدة التي تعرضت لها مجموعة المعالجة. يمكن أن يكون منهج الاختلاف في الاختلافات منهجًا منيرًا عندما تبدو لنا المعالجة غير فعالة (ارتفع معدل البطالة بعد تطبيق البرنامج عما كان قبله)، لكن المجموعة الضابطة قد وضحت لنا أن ذلك الوضع كان سيسوء أكثر في غياب ذلك التداخل.

تأثير التدريب الوظيفي على معدل البطالة في المقاطعة (أ)، بالمقارنة بالمقاطعة (ب)



تحليل الانقطاع. واحدة من طرائق خلق مجموعة معالجة ومجموعة ضابطة، هي مقارنة مخرجات مجموعةٍ ما قد تأهلت بالكاد لأن تكون تداخلًا أو مجموعة معالجة، بمخرجات مجموعة كادت أن تتأهل لتلقِّي المعالجة لكنها أخفقت بفارق طفيف للغاية. أولئك الأفراد الذين يقعون تحت قطاعٍ عشوائيٍّ ما مباشرة وأعلى قطاع عشوائي ما مباشرة، كنتائج اختبارٍ ما أو الحد الأدنى لدخل أسرةٍ ما، سيكون متطابقًا تقريبًا في العديد من الأوجه المهمة، وحقيقة أن مجموعة واحدة قد تلقَّت معالجة ولم تتلقَّها الأخرى هو أمر قد

تم بشكل عشوائي. ونتيجة لهذا نستطيع أن نقارن بين تلك المخرجات بطرائق تمنحنا نتائج ذات معنًى يتعلق بفاعلية التداخل الذي تم.

لنفترض أن مدرسةً ما فرضت على التلاميذ المتعثرين الالتحاق بصفوف صيفية. سترغب المنطقة التعليمية في أن تعرف إذا ما كان ذلك البرنامج الصيفي هو ذا قيمة أكاديمية طويلة المدى. كما هو معتاد، ستكون أي مقارنة بسيطة بين التلاميذ الذين التحقوا بالصفوف الصيفية وأولئك الذين لم يلتحقوا بها أسوأ من أن تكون غير ذات نفع. التلاميذ الذين التحقوا بتلك الصفوف قد فعلوا ذلك لأنهم متعثرون. وحتى لو كان البرنامج الدراسي الصيفي فعالًا جدًّا، فإن أولئك التلاميذ على الأرجح سوف يكون أداؤهم أسوأ على المدى الطويل من أولئك الذين لم يضطروا للالتحاق بهذه الصفوف. ما نريد معرفته هو أداء أولئك التلاميذ المتعثرين بعد أن يتلقوا تلك الصفوف الصيفية مقارنةً بالأداء الذي كانوا سيصيرون عليه لو لم يلتحقوا بها. نعم، يمكننا أن نجري نوعًا من أنواع التجارب الضابطة، والتي فيها يُنتقى تلاميذ متعثرون بشكل عشوائي كي يلتحقوا بالصفوف الصيفية أو لا يلتحقوا، ولكن ذلك سيستلزم حرمان المجموعة الضابطة من الالتحاق ببرنامج نظنه مفيدًا.

لكن ما يحدث هو أن المجموعة الضابطة ومجموعة المعالجة تُخلقان من خلال المقارنة بين أولئك التلاميذ الذين وقعوا تحت حد الصفوف الصيفية مباشرةً وأولئك الذين تجاوزوها بالكاد. فكِّر في الأمر: التلاميذ الذين رسبوا في اختبار منتصف الفصل الدراسي يختلفون كثيرًا عن زملائهم الذين لم يرسبوا فيه. لكن التلاميذ الذين حصلوا على ٥٩ بالمائة (وهي نسبة رسوب) لا يختلفون كثيرًا عن التلاميذ الذين حصلوا على ٦٠ بالمائة (وهي نسبة نجاح). إذا كان أولئك الراسبون في اختبار منتصف الفصل الدراسي يخضعون لمعالجةٍ ما، على غرار تدريس منزلي إلزامي لأجل الاختبار النهائي، فستكون لدينا مجموعة ضابطة ومجموعة معالجة، لو قارنَّا بين نتائج الاختبار النهائي لأولئك الذين كادوا أن يرسبوا في اختبار منتصف الفصل الدراسي (وتلقوا تدريسًا منزليًّا)، ونتائج أولئك الذين كادوا يرسبون في اختبار منتصف الفصل الدراسي (ولم يتلقوا تدريسًا منزليًّا).

تم استخدام هذا المنهج لتقرير فعالية الزج بالمذنبين الأحداث في السجن على سبيل الردع لأي جرائم مستقبلية قد يرتكبونها. من الواضح أن هذا النوع من التحليل لا يستطيع أن يقارن معدلات ارتداد من سُجنوا من الأحداث المذنبين إلى الجريمة، بمعدلات ارتداد الأحداث المذنبين ممن تلقوا عقوبات مخففة. الأحداث المذنبون الذين زُجَّ بهم في السجون هم بطبيعة الحال يرتكبون جرائم أكثر فداحة من الأحداث المذنبون الذين تلقوا عقوبات مخففة، فذلك هو ما أدخلهم السجون. ولا نستطيع كذلك أن نخلق مجموعة معالجة ومجموعة ضابطة بتوزيع عقوبات السجن توزيعًا عشوائيًّا (إلا إذا كنت مستعدًّا

لأن تخاطر بقضاء خمسة وعشرين عامًا في السجن عندما تنعطف يمينًا والإشارة حمراء عندما لا يكون ذلك الانعطاف مسموحًا به). استخدمت راندي هيالمارسون، التي صارت تعمل الآن باحثة بجامعة لندن، بروتوكولات معاقبة الأحداث الصارمة في ولاية واشنطن، كي تستنبط منها التأثير السببي لعقوبة السجن على السلوك الإجرامي المستقبلي. وبشكل أكثر تحديدًا، قامت بمقارنة معدلات الارتداد إلى الجريمة لأولئك الأحداث المذنبين الذين زُجَّ بهم في السجون رغم أنهم «كادوا» يُعتقون منه، بمعدلات الارتداد إلى الجريمة لأولئك الأحداث المذنبين الذين أفلتوا من السجن بعد أن «كادوا» يُزجُّون فيه (ويكون هذا عادة عن طريق دفع غرامة أو الخضوع للمراقبة الشرطية)[٩]

يقوم النظام القضائي في واشنطن ببناء مخطط شبكي لكل مجرم مُدان، ويستخدم هذا المخطط في تقرير عقوبته. يقيس محور x سوابق المجرم التي ثبتت عليه وحُكم بها. فمثلًا تمثل كل جناية نقطة واحدة في الرسم البياني، وتمثل كل جنحة ربع نقطة. يتم تقريب مجموع النقاط إلى عدد صحيح (وهو الأمر الذي سرعان ما ستتضح لنا أهميته). وفي نفس الوقت يقيس المحور y فداحة الجرم على مقياس يتدرج من E (الأقل فداحة) إلى A+ (الأكثر فداحة). يتم حساب عقوبة الحدث المُدان بشكل حرفي عن طريق إيجاد المربع المناسب له في المخطط الشبكي: فالمذنب الذي حاز ما يعادل النقطتين من الجرائم السابقة، والذي ارتكب جناية من الدرجة B، سوف يعاقب بقضاء ما بين خمسة عشر شهرًا وستة وثلاثين شهرًا في سجن الأحداث. أما المذنب المُدان الذي له سوابق تعادل نقطة واحدة، والذي ارتكب جناية من نفس الدرجة، لن يُزَجَّ به في السجن. كان الانقطاع هو ما أوجد الدافع لهذه الاستراتيجية البحثية. قامت هيالمارسون بمقارنة مخرجات المذنبين المُدانين التي وقعت أعلى حد عقوبة السجن وتحته مباشرة. تصف ذلك في البحث بقولها: «لو كان هناك شخصان ارتكبا جريمتين حديثتين من الدرجة C+، ولهما سوابق جنائية قدرها ٢¾ و٣، فحينها سيسجن الشخص الأخير فقط في سجن الولاية».

ولغرض البحث يتساوى هذان الشخصان بشكل جوهري إلى أن يُزَجَّ بأحدهما في السجن. وعند تلك اللحظة تتباين سلوكياتهما تباينًا حادًّا. المذنبون الأحداث الذين سُجنوا هم أقل احتمالًا لأن يدانوا بجريمة أخرى (بعد أن يطلق سراحهم من السجن).

إننا نُعنى بكل ما هو ناجح. وهي الحقيقة التي تصدق على الطب والاقتصاد والمشاريع والعدالة الجنائية – وعلى كل شيء. لكن السببية أشبه بجوزة صلبة القشرة من الصعب كسرها، حتى في الحالات التي يكون فيها السبب والتأثير واضحَين وضوح الشمس. لفهم التأثير الحقيقي للمعالجة نحتاج لأن نعرف «التأثير المعاكس»، وهو الذي ما كان سيحدث إذا غابت المعالجة أو التداخل. عادةً ما يكون هذا التأثير المعاكس صعبَ الملاحظة أو مستحيلَها. تَأَمَّلْ هذا المثال غير الإحصائي: هل جعل غزوُ العراق الولاياتِ المتحدةَ أكثرَ أمنًا؟

هناك إجابة وحيدة أمينة علميًّا، وهي أننا لن نعرف أبدًا. والسبب في أننا لن نعرف أبدًا هو أننا لا نعرف – ولا نستطيع أن نعرف – ما كان سيحدث لو أن الولايات المتحدة الأمريكية لم تَجْتَحِ العراق. صحيح أن الولايات المتحدة لم تعثر على أي أسلحة دمار شامل هناك. لكنه من الممكن جدًّا أن يكون صدام حسين حدث وأنه في اليوم الذي تلا عدم اجتياح الولايات المتحدة للعراق، وهو في مسبح حمامه، قد خطر بباله خاطر فقال لنفسه: «إنني فعلًا بحاجة إلى قنبلة هيدروجينية، يا ترى أسيبيعني الكوريون الشماليون واحدة منها؟» وبعدها، من يعرف ما كان سيحدث؟

وبالطبع كان من الممكن أن يدلف صدام حسين إلى مسبح حمامه نفسه في نفس اليوم الذي تلا عدم اجتياح الولايات المتحدة للعراق، ثم قال لنفسه: «إنني فعلًا بحاجة إلى ــــــــ»، وهنا انزلقت قدماه على قطعة الصابون، فارتطم رأسه بالتمثال الرخامي الصغير الذي يزين المسبح فمات. في تلك الحالة كان العالم سيتخلص من صدام حسين من دون تلك التكاليف الهائلة التي تكبَّدتها الولايات المتحدة في الاجتياح. من يدري ما كان سيحدث؟

الغرض من أي تقييم لبرنامج هو تقديم نوع من أنواع التأثير المعاكس في مقابل المعالجة أو التداخل الذي يمكن قياسه. في حالة التجارب الضابطة ذات الانتقاء العشوائي، تصير المجموعة الضابطة هي التأثير المعاكس. وفي حالة كون التجربة الضابطة غير عملية أو غير أخلاقية، نحتاج حينها لأن نجد طريقة أخرى لمقاربة هذا التأثير العكسي. وفهمُنا للعالم يعتمد على إيجادنا لطرائق بارعة للقيام بهذا الأمر.

الخاتمة

خمسة أسئلة يساعدنا الإحصاء في الإجابة عنها

في زمن ليس ببعيد كان جمع المعلومات أكثر صعوبة، وكان تحليلها أبهظ ثمنًا. تخيل أنك تدرس معلومات عن مليون معاملة بطاقة ائتمان في حقبة – منذ عقود قليلة مضت – كانت كل الإيصالات فيها ورقية، ولم تتواجد فيها الحواسيب التي تحلل تلك البيانات المتراكمة. خلال الكساد الكبير، لم تتواجد إحصائيات رسمية يقاس من خلالها مدى فداحة المشكلات الاقتصادية. لم تجمع الحكومة معلومات رسمية عن إجمالي الناتج المحلي أو البطالة، وهو ما يعني أن السياسيين وقتها كانوا يتعاملون مع الاقتصاد كمن يتوغل في غابة من دون بوصلة. ففي عام ١٩٣٠ أعلن هربرت هوفر انتهاء الكساد الكبير استنادًا إلى البيانات المنتهية الصلاحية غير الدقيقة التي أتيحت له وقتها. وفي خطاب حالة الاتحاد قال للأمريكيين إن مليونين ونصف المليون منهم يعانون البطالة. لكن الواقع أن خمسة ملايين أمريكي كانوا بلا وظائف، وكان معدل البطالة يزيد مائة ألف كل أسبوع. أو كما أشار جيمس سورويكي مؤخرًا في مجلة النيويوركر قائلًا: «كانت واشنطن تصنع سياساتها في الظلام»[1]

لكن الآن صارت البيانات تغمرنا. وذلك أمر جيد في معظمه، فتلك الأدوات الإحصائية التي ذكرت في هذا الكتاب يمكن أن تستخدم للتعامل مع بعض أهم التحديات الاجتماعية التي تواجهنا. وفي هذا الإطار وجدتُ أنَّ الأنسب أنْ أُنهي هذا الكتاب بأسئلة لا بإجابات. فبينما نحاول أن نهضم ونحلل تلك الكميات الهائلة من المعلومات، إليك خمسة أسئلة مهمة (وأُقرُّ أنها عشوائية) سوف تنطوي إجاباتها الدالة اجتماعيًّا على الكثير من الأدوات التي ذكرناها في هذا الكتاب.

ما مستقبل كرة القدم الأمريكية؟

في عام ٢٠٠٩ طرح مالكولم جلادويل في مقاله بمجلة نيويوركر سؤالًا بدا لي للوهلة الأولى سؤالًا إثاريًّا استفزازيًّا من دون داعٍ، وهو: فيم تختلف مصارعة الكلاب عن كرة القدم الأمريكية؟[2] ذلك الرابط بين النشاطين قد نبع من حقيقة أن مايكل فيك، لاعب مركز

خط الوسط، والذي عوقب بالسجن لتورطه في إدارة حلبة لمصارعة الكلاب، قد أعيد إدراجه في الاتحاد الوطني لكرة القدم الأمريكية في نفس الوقت الذي بدأت تتواتر فيه معلومات عن أن الإصابات التي تتسبب فيها ممارسة اللعبة، ترتبط بالاكتئاب وفقدان الذاكرة والخرف، وغيرها من المشكلات العصبية التي يصاب بها المرء في سِنِي عمره المتقدمة. كانت الفرضية الأساسية التي وضعها جلادويل هي أن كرة القدم الأمريكية الاحترافية ومصارعة الكلاب كلتيهما نشاطان يدمران المشاركين فيه. ومع وصولي لنهاية المقال صِرت مقتنعًا بأنه قد طرح نقطة مثيرة للاهتمام.

إليك ما نعرفه. هناك أدلة متراكمة على أن الارتجاجات المخية، وغيرها من الإصابات الدماغية التي تنطوي عليها ممارسة كرة القدم الأمريكية، قادرة على إحداث أضرار عصبية خطيرة ودائمة. (لوحظت ظواهر مشابهة لدى لاعبي الملاكمة والهوكي). وكان كثير من أبرز لاعبي الاتحاد الوطني لكرة القدم الأمريكية السابقين قد حكوا علنًا عن الصراعات التي خاضوها بعد اعتزالهم مع الاكتئاب وفقدان الذاكرة والخرف. وربما كانت أكثر هذه القصص مأساوية هي قصة ديف دويرسون، اللاعب السابق بمركز الدفاع والفائز ببطولة السوبر مع فريق شيكاجو بيرز، والذي أقدم على الانتحار بإطلاق النار على صدره بعد أن ترك رسالة بها تعليمات مفصلة لعائلته، طالبًا منهم أن يطلبوا من الباحثين دراسة مخه بعد وفاته.

في استقصاء أجري عبر الهاتف لألف لاعب سابق من لاعبي الاتحاد الوطني لكرة القدم الأمريكية، والذين تم انتقاؤهم عشوائيًّا من بين من لعبوا لثلاثة أعوام على الأقل في الاتحاد، أفاد ما نسبته ٦,١ بالمائة من الذين تجاوزوا الخمسين عمرًا من بينهم، أنهم قد شخص الأطباء لديهم حالات من «الخرف والألزهايمر وغيرها من الأمراض التي ترتبط بالذاكرة». وهو المعدل الذي يعادل خمسة أمثال المعدل القومي لهذه الفئة العمرية. وبالنسبة إلى اللاعبين الأصغر سنًّا، جاء معدل إصابتهم بتلك الأمراض زائدًا تسعة عشر مِثْلًا عن المعدل القومي. ولقد قام المئات من لاعبي الاتحاد الوطني السابقين بمقاضاة الاتحاد وصانعي خوذات كرة القدم كليهما، زاعمين أنهم قد أخفوا معلومات عن مخاطر إصابات الرأس[٢].

كانت آن ماكي واحدة من الباحثين الذين درسوا تأثيرات إصابات الرأس، وهي تدير معمل أمراض الأعصاب بمستشفى قدامى المحاربين في بيدفورد بولاية ماساتشوستس. (وتصادف أنها من قامت بالجزء المتعلق بعلم الأعصاب بدراسة فارمنجهام لأمراض القلب). وثَّقت الدكتورة ماكي حدوث تراكم في بروتين غير معتاد يسمى بروتين التاو في أمخاخ الرياضيين الذين يعانون من إصابات بالدماغ، كالملاكمين ولاعبي كرة القدم الأمريكية. يؤدي ذلك إلى حالة تسمى باعتلال الدماغ الرضحي المزمن، والذي يرمز

له بالأحرف CTE، والذي هو اضطراب عصبي استفحالي له الكثير من نفس علامات الألزهايمر.

وفي نفس الوقت كان هناك باحثون آخرون يوثِّقون الارتباط بين كرة القدم الأمريكية والإضرار بالدماغ. فقد قام كيفن جوسكيويتز، مدير برنامج البحث في الارتجاجات المخية الناتجة عن ممارسة الرياضة بجامعة نورث كارولينا، بتركيب مستشعرات داخل خوذات لاعبي فريق نورث كارولينا لكرة القدم الأمريكية؛ بهدف تسجيل قوة الضربات التي تصيب الدماغ وطبيعتها. أفادت بياناته بأن اللاعبين يتعرضون بشكل روتيني لضربات على الرأس، تعادل الواحدة منها ارتطام رأس شخص بالزجاج الأمامي لسيارة تنطلق بسرعة خمسة وعشرين ميلًا في الساعة.

وإليك ما لا نعرفه. هل تلك الدلائل التي تكشفت حتى الآن تعتبر ممثلة للمخاطر العصبية الطويلة الأمد التي يواجهها جميع لاعبي كرة القدم الأمريكية المحترفين؟ أم هي مجرد «تجمُّع» لمخرجات شاذة هي عبارة عن انحراف إحصائي؟ وحتى لو اتضح أن لاعبي كرة القدم الأمريكية معرَّضون بالفعل لأن يصابوا باضطرابات عصبية مع تقدُّمهم في العمر، فسنظل مضطرين لأن ندرس تلك السببية. هل طبيعة أولئك الرجال الذين يلعبون كرة القدم الأمريكية (والملاكمة والهوكي) تكون هي المسئولة عن تعريضهم لهذه المشكلة؟ هل من الممكن أن تساهم عوامل أخرى، كتعاطي مركبات الستيرويد في إحداث هذه المشكلات العصبية في المستقبل؟

لو أشارت الأدلة المتراكمة إلى وجود رابط سببي واضح بين ممارسة كرة القدم الأمريكية وحدوث أذًى طويل المدى بالدماغ، فحينها سيطل سؤال ملحٌّ برأسه على اللاعبين (وذوي اللاعبين الأصغر سنًّا) والمدربين والمحامين ومسئولي الاتحاد الوطني لكرة القدم الأمريكية، بل وربما المشرعين الحكوميين، ألا وهو: أهناك من طريقة أخرى لِلَعب كرة القدم تقلل مخاطر إصابات الرأس أو تَحُول دونها تمامًا؟ وإذا لم يكن من سبيل، فماذا إذن؟ كان ذلك هو الذي قصده مالكولم جلادويل من مقارنته كرة القدم الأمريكية بمصارعة الكلاب. فقد بيَّن أن العامة يُبغضون تصارع الكلاب؛ لأن مالك الكلب يُخضع كلبه عن عمد لمسابقة تنطوي على معاناة وموت. ويتابع متسائلًا: «ولأي سبب؟ لأجل الترفيه عن جمهور ما، ولأجل راتب يتقاضاه اللاعبون عن كل يوم يلعبون فيه. كانت مصارعة الكلاب رياضة تلقى قبولًا واسعًا لدى الجماهير الأمريكية خلال القرن التاسع عشر. لكننا الآن لم نعُد نجدها رياضة مقبولة أخلاقيًّا».

حاليًّا صارت كل أنواع التحليلات الإحصائية التي أوردناها في هذا الكتاب تقريبًا تُستخدم لتقرير إذا ما كانت كرة القدم الأمريكية التي نعرفها ينتظرها مستقبل أو لا.

ما الذي يسبب ذلك الارتفاع الدراماتيكي في حالات التوحد؟ (لو كان هناك بالفعل شيء يسبب هذا).

في عام ٢٠١٢ أعلنت مراكز التحكم في الأمراض أن ١ من كل ٨٨ طفلًا أمريكيًّا شُخصت لديهم حالات اضطراب الطيف التوحدي (استنادًا إلى بيانات جُمعت عام ٢٠٠٨)[٤]. ارتفع معدل تشخيص مرض التوحد لدى الأطفال من طفل واحد من كل ١١٠ أطفال عام ٢٠٠٦، وطفل واحد من كل ١٥٠ طفلًا عام ٢٠٠٢ – وهو ما يعني أن العدد قد تضاعف في أقل من عشر سنين. واضطرابات الطيف التوحدي، والتي يُرمز لها بالأحرف (ASDs)، هي مجموعة من إعاقات النمو، والتي تصاحبها إعاقات في التنشئة الاجتماعية والتواصل والسلوك. ومصطلح «الطيف» هنا يشير إلى أن التوحد يجمع نطاقًا واسعًا من الحالات المعرفة سلوكيًّا[٥] والأولاد الذكور هم أكثر عرضة للإصابة باضطرابات الطيف التوحدي من الإناث بمعدل قدره خمسة أمثال (وهو ما يعني أن معدل الإصابة لدى الذكور يزيد حتى على معدل ١ من كل ٨٨ طفلًا).

نجد هنا أول الأسئلة المثيرة لاهتمامنا من الناحية الإحصائية، وهو أنحن نشهد وباءَ «توحدٍ»، أم «وباءً تشخيصيًّا» أم خليطًا من الاثنين؟[٦] فخلال العقود الماضية كان الأطفال الذين يعانون طيف التوحد تظهر عليهم أعراض لا تشخَّص عادة، أو كانت صعوبات النمو التي يمرون بها توصَّف عامة على أنها «إعاقات في التعلم». لكن صار الأطباء والآباء والمعلمون على وعي أكبر بأعراض اضطرابات طيف التوحد، وهو الأمر الذي يؤدي إلى مزيد من تشخيصات هذا المرض، بصرف النظر عما إذا كان معدل الإصابة به في زيادة أو لم يكن.

أيًّا كان الوضع، فإن الارتفاع الصادم في معدل الإصابة باضطرابات طيف التوحد يمثل تحديًا للعائلات والمدارس وبقية المجتمع. إذ يبلغ متوسط كلفة التعامل مع اضطراب طيف التوحد على مدى حياة المريض كلها ٣,٥ ملايين دولار[٧] ورغم أنه قد صار من الواضح كون الأمر قد تحوَّل إلى وباء، فإننا لا نعرف إلا أقل القليل عن مسببات هذا الأمر. قال توماس إنسيل مدير المعهد الوطني للصحة العقلية: «أهي الهواتف المحمولة؟ أم الموجات فوق الصوتية؟ أم المياه الغازية الخاصة بالحمية؟ لكل واحد من الآباء نظريته الخاصة. حتى هذه اللحظة نحن بكل بساطة لا نعرف»[٨].

ما المختلف والمتفرد في حيوات الأطفال المصابين باضطراب طيف التوحد وخلفياتهم؟ ما أهم الاختلافات الفسيولوجية بين الأطفال المصابين باضطراب طيف التوحد وغيرهم؟ هل يختلف معدل الإصابة بهذا المرض من دولة لأخرى؟ لو كان ذلك هو الحال، فلماذا؟ هنا تقدِّم لنا أعمال التحريات الإحصائية التقليدية بعض الدلائل.

واحدة من الدراسات الحديثة التي أجراها باحثون بجامعة كاليفورنيا بمدينة دافيس حددت عشر مواقع بكاليفورنيا تبلغ فيها معدلات الإصابة بالتوحد ضعف نظائرها في المناطق المحيطة بها، وكل بؤر التوحد تلك هي أحياء يتركز فيها آباء من بيض البشرة ذوي مستويات عالية من التعليم[٩]. أهذا دليل أم إنه مصادفة؟ أم إنه يعكس حقيقة مفادها أن الأسر الميسورة نسبيًّا هي الأكثر قابلية للإصابة باضطراب طيف التوحد؟ نفس هؤلاء الباحثين يقومون أيضًا بإجراء دراسة، يجمعون فيها عينات غبار من منازل ١٣٠٠ أسرة من التي تضم طفلًا متوحدًا؛ لاختبار ما بها من مواد كيماوية أو غيرها من الملوثات البيئية التي قد تكون قد لعبت دورًا سببيًّا في الأمر.

في نفس الوقت كان باحثون آخرون قد فطنوا إلى ما بدا أنه مكوِّن جيني للتوحد من خلال دراسة اضطراب طيف التوحد بين التوائم المتطابقة وغير المتطابقة[١٠] فقد وجد أن أرجحية إصابة طفلين من نفس الأسرة باضطراب طيف التوحد أعلى بين التوائم المتطابقة (الذين يشتركون في نفس التركيب الجيني)، منها بين التوائم غير المتطابقة (حيث لا يختلف التشابه الجيني بين التوأمين عنه لدى أي شقيقين غير توأمين). لم تنفِ هذه النتيجة العوامل البيئية المهمة، بل وحتى التفاعل بين العوامل البيئية والجينية. فصحيح أن للإصابة بمرض القلب مكوِّنًا جينيًّا مهمًّا، لكن من الجلي أيضًا أن التدخين والحمية الغذائية والتريض، وغيرها من العوامل السلوكية والبيئية الكثيرة، هي مكونات مهمة أيضًا.

واحدة من أهم مساهمات التحليل الإحصائي التي حققها حتى الآن هي فضح الأسباب الزائفة، وهي الأسباب التي نشأت من الخلط بين الارتباط والسببية. عادة ما يَظهر اضطراب طيف التوحد فجأة بين عيد ميلاد الطفل الأول وعيده الثاني. ولقد أدى هذا إلى اعتقاد شائع بأن اللقاحات التي يتلقاها الطفل – وخاصة ذلك اللقاح الثلاثي الذي يقي الحصبة والنكاف والحصبة الألمانية، والذي يرمز له بأحرف (MMR) – هي التي تسبب الإصابة بالتوحد. حتى إن دان برتون، عضو الكونجرس عن ولاية إنديانا، قد صرحَّ لصحيفة نيويورك تايمز قائلًا: «حُقن حفيدي بتسعة لقاحات في يوم واحد، وكانت سبعة منها تحتوي على ثيمروسال، والذي يكون الزئبق نسبة ٥٠ بالمائة منه كما تعلم، ولم يمضِ وقت طويل بعدها حتى صار طفلًا متوحدًا»[١١]

كان العلماء قد فندوا ذلك الارتباط الزائف بين الثيمروسال واضطراب طيف التوحد تفنيدًا قاطعًا. فلم تتراجع معدلات الإصابة بالتوحد بعد أن أزيل مركب الثيمروسال من اللقاح الثلاثي، ولم تنخفض أيضًا في الدول التي لا تستخدم اللقاح من الأساس. لكن رغم ذلك يظل ذلك الربط الزائف قائمًا لدرجة أن بعض الآباء يرفضون أن يتلقى أطفالهم هذا اللقاح. والمفارقة هنا أن هذا لا يقي الأطفال الإصابة بالتوحد، في حين أنه يعرض الأطفال لمخاطر الإصابة بأمراض خطيرة أخرى (والمساهمة في نشر هذه الأمراض بين السكان).

يعتبر التوحد واحدًا من أكبر التحديات الطبية والاجتماعية في عصرنا هذا. وإننا لا ندرك عنه إلا أقل القليل مقارنة بتأثيره العظيم (والآخذ في التعاظم) على سلامتنا الجمعية. يستخدم الباحثون كل ما ورد في هذا الكتاب (وكثير غيرها) لتغيير هذا الواقع.

كيف لنا أن نحدد المدرسين الأكْفاء والمدارس الممتازة ومن ثم مكافأتهم؟

إننا بحاجة لمدارس ممتازة. ونريد مدرسين أكْفاءً كي نحظى بتلك المدارس الممتازة. ومن ثم فإن هذا يستلزم، منطقيًّا، أن نكافئ المدرسين الأكْفاء والمدارس الممتازة، وأن نقيل المدرسين غير الأكْفاء ونغلق المدارس السيئة.

كيف لنا أن نفعل ذلك بالضبط؟

تمنحنا نتائج الاختبارات مقياسًا موضوعيًّا لأداء التلاميذ. لكننا نعلم أن من بين التلاميذ من سيُبلون بلاءً أفضل بكثير في الاختبارات الموحدة من غيرهم، لأسباب لا علاقة لها بما يحدث في داخل قاعة الدرس أو في المدرسة. وقد يبدو الحل البسيط هنا هو تقييم المدارس وما فيها من مدرسين استنادًا إلى مدى التقدم الذي يحرزه تلاميذهم على مدى فترة زمنية محددة. ماذا تَعلَّم التلاميذ بعدما التحقوا بصف معين مع مدرس معين؟ وماذا تعلموا بعد الالتحاق بعام؟ والفارق بين الاثنين هو «القيمة المضافة» لهذا الصف.

بل إننا نستطيع أن نستخدم الإحصاء لتنقيح إدراكنا لهذه القيمة المضافة، عن طريق اعتبار الخصائص الديمغرافية لتلاميذ قاعة درس معينة، كالعِرق ومستوى الدخل وأدائهم في اختبارات أخرى (والذي يعتبر مقياسًا للموهبة). فإذا ما حقق معلم أو معلمة ما تقدمًا كبيرًا مع تلاميذ كانوا متعثرين في السابق، فيمكننا أن نصف أيًّا منهم بأنه معلم كفء.

ها قد وجدناها. نستطيع الآن أن نقيِّم كفاءة المعلم بدقة إحصائية. وبالطبع، فالمدارس الممتازة هي المليئة بالمعلمين الأكْفاء.

كيف تتفعل هذه التقييمات الإحصائية بشكل عملي؟ في عام ٢٠١٢ غامرت مدينة نيويورك وقامت بإعلان ترتيب ١٨٠٠٠ معلم بالمدارس الحكومية طبقًا «لتقدير القيمة المضافة» الذي يقيس التقدم في نتائج اختبارات تلاميذهم، مع أخذ خصائص متعددة تميز هؤلاء التلاميذ في اعتبارهم[١٢]. وكانت صحيفة لوس آنجليس تايمز قد نشرت ترتيبًا مشابهًا لمعلمي لوس آنجيليس عام ٢٠١٠.

ولقد جاءت ردود الأفعال في نيويورك ولوس آنجيليس كلتيهما صاخبة متضاربة. كان آرنى دانكان وزير التعليم الأمريكي داعمًا لمثل هذه التقييمات التي تستند إلى القيمة المضافة بشكل عام؛ فهي تقدِّم معلومات جديدة لم تتواجد قبلها. وبعد أن نشرت لوس آنجيليس بياناتها، صرح دانكان لصحيفة نيويورك تايمز قائلًا: «لم يعد الصمت خيارًا ممكنًا». وقدمت إدارة أوباما حوافز مالية للولايات التي وضعت مؤشرات قائمة على القيمة

المضافة، تقوم عليها مكافأة المعلمين وترقيتهم. يشير مؤيدو مقاييس التقييم تلك إلى أنها تحسُّن كبير عن الأنظمة السابقة التي كان المعلمون يتقاضون في ظلها رواتب موحدة لا تُقِيم للأداء في الصف الدراسي أي وزن.

لكن في المقابل حذَّر خبراء كثيرون من أن هذه الأنواع من تقييمات المعلمين تنطوي على هوامش خطأ كبيرة، ومن الممكن أن تخرج بنتائج مضللة. بل إن الاتحاد الممثل لمعلمي نيويورك قد أنفق ما يزيد عن ١٠٠٠٠٠ دولار على حملة دعاية في الصحف تقوم على العنوان الصحفي «ليست هذه طريقة لتقييم المعلمين»[١٣] يزعم المعارضون أن تقييمات القيمة المضافة تخلق دقة زائفة، سيُسيء الآباء والمسئولون الحكوميون الذين لا يفهمون حدود هذه التقييمات استخدامَها.

تبدو هذه حالة من الحالات التي يصير الجميع فيها على حق – حتى نقطة معينة. حذَّر دوج ستايجر، عالم الاقتصاد بكلية دارتموث والذي يبحث كثيرًا في بيانات القيمة المضافة الخاصة بالمعلمين، من أن هذه البيانات «صاخبة» بطبعها. فنتائج معلمٍ ما عادةً ما تعتمد على اختبار واحد أقيم في يوم واحد عن طريق مجموعة واحدة من التلاميذ. فعواملُ مِن كل صنف ونوع يمكن أن تسبب تذبذبات عشوائية – عوامل على غرار أن تكون مجموعة التلاميذ صعبة بطبعها أو يكون جهاز تكييف الهواء قد تعطل في يوم الاختبار. يبلغ ارتباط الأداء لمعلم واحد من عام لآخر، والذي يستخدم هذه المؤشرات، ٣٥. (المثير للاهتمام هنا أن ارتباط الأداء للاعبي البطولة الكبرى للبيسبول من عام لآخر يبلغ أيضًا نحو ٣٥. عندما يقاس عن طريق معدل ضرب الكرات للضاربين ومتوسط الجريات المستحقة للرامين)[١٤]

يقول ستايجر إن البيانات المتعلقة بكفاءة المعلمين مفيدة، لكن ما هي إلا أداة واحدة من الأدوات التي تتضمنها عملية تقييم أداء المعلمين. تغدو البيانات «أقل صخبًا» عندما تتاح للسلطات الرسمية بيانات لأعوام أكثر قضاها معلم معين يدرِّس صفوفًا مختلفة (وهو الأمر الذي ينطبق كذلك على الرياضيين عندما تتوافر بيانات أكثر عن مباريات أكثر لعبوها ومواسم أكثر قضوها). في حالة تقييم معلمي نيويورك تم تدريب مديري المدارس المدرجين في تلك المنظومة على الاستخدام الأنسب لبيانات القيمة المضافة ومواطن القصور التي تنطوي عليها. لكن العامة لم يتلقوا ذلك التدريب، ونتيجة لذلك صارت تقييمات المعلمين تُصوَّر عادة على أنها مؤشر أكيد على «الجيد» من المعلمين و«السيء» منهم. وإننا نحب الترتيب – كما في ترتيب الكليات الذي تصدره شبكة يو إس آند ورلد ريبورت – حتى عندما لا تدعمه البيانات.

ثم يقدم ستايجر تحذيرًا أخيرًا من نوع مختلف، ألا وهو أننا يجدر بنا أن نتيقن من كون المخرجات التي نقيسها، مثل نتائج اختبار موحد، ستتماشى على المدى الطويل

مع ما نهتم به. فلم يكن من المدهش أن بعض البيانات المتفردة الصادرة عن أكاديمية القوات الجوية قد بيَّنَت أن نتائج الاختبارات التي تبدو برَّاقة الآن لن تصير ذهبًا في المستقبل. فأكاديمية القوات الجوية، كغيرها من الأكاديميات العسكرية، تقوم بتوزيع عشوائي لطلبتها على صفوف مختلفة من مقررات موحدة، كمقرر مقدمة في حساب التفاضل والتكامل. وهذا التوزيع العشوائي من شأنه أن ينفي أي تأثير انتقائي محتمل عند المقارنة بين الأساتذة من حيث الكفاءة، وبمرور الوقت نستطيع أن نفترض أن جميع الأساتذة يحظون بطلبة يتمتعون بمواهب متشابهة (على النقيض من معظم الجامعات التي يستطيع فيها الطلاب ذوو القدرات المتباينة أن ينتقوا مقررات دراسية مختلفة). كذلك تستخدم أكاديمية القوات الجوية نفس المناهج والاختبارات في كل صف من صفوف نفس المقرر. استغل كلٌّ من سكوت كاريل وجيمس ويست، الأستاذان بجامعة كاليفورنيا بمدينة دافيس وكذلك في أكاديمية القوات الجوية، هذا النظام الأنيق في الإجابة على واحد من أهم الأسئلة المطروحة في مجال التعليم العالي، ألا وهو: أي الأساتذة الأكثر كفاءة؟[١٥]

والإجابة هي: هم الأساتذة الأقل خبرة والذين يحملون درجات علمية من جامعات عريقة أقل من غيرهم. أولئك الأساتذة يحظون بطلبة عادة ما يُبلون بلاءً أفضل في الاختبارات الموحدة للمقررات التقديمية. وهم كذلك يحصلون على تقييمات أفضل من تلاميذهم في مقرراتهم تلك. من الطبيعي أن يكون أولئك المعلمون الشبان المتحمسون أكثر إخلاصًا للتدريس من الأساتذة الهرمين العصبيين الذين يحملون درجات دكتوراه من جامعات كَهارفارد. لا بد أن أولئك الرفاق المسنين ما زالوا يستخدمون وريقات التدريس المصفرَّة التي كانوا يستخدمونها عام ١٩٧٨، ولا بد أنهم يظنون برنامج الباور بوينت نوعًا من أنواع مشروبات الطاقة، عدا أنهم لا يعرفون ما هي مشروبات الطاقة من الأساس. من الواضح أن البيانات تحثُّنا على أن نُقيل أولئك المتحجرين، أو على الأقل نتركهم يتقاعدون بكرامة.

لكن لنتريث ولا نَقُمْ بإقالة أحدًا بعد. فإن دراسة أكاديمية القوات الجوية قد خلصت إلى نتيجة أخرى مهمة تتعلق بأداء التلاميذ على المدى الأطول. وجد كاريل وَويست أنه في المقررات التقديمية لمادتَي الرياضيات والعلوم، كان الطلبة الذين حظوا بأساتذة أكبر خبرة (وأكثر حيازة لدرجات علمية) يُبلون بلاءً أفضل في المقررات الإلزامية اللاحقة، من زملائهم الذين درَّس لهم أساتذة أقل خبرة في المقررات التقديمية. أحد التفسيرات المنطقية لهذا الأمر هو أن الأساتذة الأقل خبرة يميلون أكثر لأن «يدرِّسوا التلاميذ لأجل الاختبار» في تلك المقررات التقديمية. وهو الأمر الذي تتمخض عنه نتائج اختبارات مثيرة للإعجاب، وطلبة سعداء يَملئون استمارات تقييم معلميهم.

في الوقت نفسه نجد الأساتذة الهرمين العصبيين (الذين كدنا نقيلهم من العمل في

الفقرة قبل الماضية) يولون تركيزًا أقل إلى الاختبار، ويركزون أكثر على المفاهيم المهمة، والتي تخدم المقررات اللاحقة وتفيد في حياةٍ ما بعد التخرج من أكاديمية القوات الجوية.

إننا نحتاج بالطبع لأن نقيِّم معلمي المدارس وأساتذة الجامعات. لكن علينا أن نتأكد من فعل ذلك بالطريقة الصحيحة. والتحدي الذي يواجه السياسة الطويلة المدى، وهو التحدي المتجذر في الإحصاء، هو كيفية وضع نظام يكافئ المعلم على القيمة التي يضيفها في قاعة الدرس.

ما أفضل الأدوات لمكافحة الفقر في العالم؟

إن معرفتنا بكيفية تقليل فقر البلاد الفقيرة ضئيلةٌ بشكل صادم. صحيح أننا ندرك الأمور التي تميز البلدان الغنية عن الفقيرة، كمستويات التعليم وكفاءة الحكومات. وصحيح أيضًا أننا شاهدنا دولًا، مثل الهند والصين، تحوِّل نفسها من الناحية الاقتصادية على مدى العقود الماضية. لكن حتى مع تلك المعرفة، لا يتضح لنا ما الخطوات التي ينبغي أن نتخذها لتقليل الفقر في بلدان مثل مالي أو بوركينا فاسو. من أين يجدر بنا أن نبدأ؟

تقوم عالمة الاقتصاد الفرنسية إستير دوفلو بتعديل معرفتنا بالفقر العالمي من خلال تعديل إحدى الأدوات القديمة تعديلًا تحديثيًّا ليلائم أغراضًا جديدة، وهذه الأداة هي: التجربة الضابطة ذات الانتقاء العشوائي. تقوم دوفلو، التي تدرِّس في معهد ماساتشوستس للتكنولوجيا، حرفيًّا بإجراء تجارب على تداخلات مختلفة لتحسين حيوات الفقراء في الدول النامية. فمثلًا يعتبر تغيُّب المعلمين عن العمل واحدة من المشكلات المزمنة في مدارس الهند، وخاصة في مدارس المناطق الريفية التي لا يعمل بها سوى معلم واحد. قامت دوفلو وريما هانا المؤلفة المشاركة معها باختبار حل حاذق قائم على التكنولوجيا على عينة عشوائية مؤلفة من ستين مدرسة في ولاية راجستان الهندية[١٦]، لا يعمل بكل منها سوى معلم واحد. تم منح معلمي تلك المدارس التجريبية الستين مكافآت للمواظبة على الحضور. لكن الجزء الإبداعي من الحل هو أن المعلمين قد أُعطوا كاميرات تطبع الوقت والتاريخ على الصور ولا سبيل للعبث بها. وكل يوم كانوا يُثبتون حضورهم بأن تلتقط صورة لهم مع تلاميذهم[١٧]

انخفض معدل التغيب عن العمل إلى النصف بين معلمي المدارس التجريبية، مقارنةً بمعلمي مدارس المجموعة الضابطة الستين الذين تم انتقاؤهم عشوائيًّا. وتحسنت نتائج اختبارات التلاميذ، وانتقل الكثير منهم إلى المستويات الأعلى (وأُراهن أن تلك الصور كانت رائعة أيضًا).

وتضمنت واحدة من التجارب التي أجْرَتها دوفلو في كينيا منْح مجموعة منتقاة عشوائيًّا من المزارعين دعمًا ماليًّا بسيطًا كي يبتاعوا به أسمدة زراعية بعد الحصاد. كانت

أدلة سابقة قد أظهرت أن الأسمدة ترفع من عائد المحصول بشكل كبير. كان المزارعون على علم بهذا الأمر، لكنهم كانوا وقت غرس المحصول الجديد يفتقرون إلى المال اللازم لشراء تلك الأسمدة. يساهم هذا في استدامة ما يسمى باسم «مصيدة الفقر»؛ حيث يكون مزارعو الكفاف أفقر من أن يجعلوا أنفسهم أقل فقرًا. وجدت دوفلو ومؤلفوها المشاركين أن دعمًا ماليًّا صغيرًا – توفير الأسمدة مجانًا – يقدَّم للمزارعين، بينما هم ما زالوا يحتفظون بالمال عقب الحصاد، قد زاد من استخدام الأسمدة بنسبة تتراوح ما بين ١٠ و٢٠ بالمائة، مقارنةً باستخدامها في المجموعة الضابطة[١٨]

بل إن إستير دوفلو قد انخرطت في معمعة حرب الجنسين. من الأكثر مسئولية عندما يتعلق الأمر بإدارة أمور الأسرة المالية؛ أهُم الرجال أم النساء؟ في البلدان الثرية يكون هذا واحدًا من الأمور التي يتنازع الزوجان عليها في جلسات الاستشارات الزوجية. أما في البلدان الفقيرة، فهو يحدد، حرفيًّا، إذا ما كان الأطفال في الأسرة سيجدون ما يكفي من الطعام. تزعم أدلة قادمة من الحكايات الشعبية أنه منذ فجر الحضارة كانت النساء تعطين الأولوية لصحة أولادهن ورفاهتهم، بينما يميل الرجال لأن ينفقوا أجورهم كاملة على معاقرة الخمر بالحانات المحلية (أو ما كان يناظرها لدى رجال الكهوف). هذه الأدلة الشعبية لا تفعل في أسوأ حالاتها إلا أنها تدعم تلك الأفكار المقولبة القديمة قدم التاريخ. أما في أفضل حالاتها تكون عصية على الإثبات؛ وذلك لأن ماليات الأسرة تتشابك إلى حد ما. كيف لنا أن نفرق بين الطرق التي يختار الأزواج والزوجات أن ينفقوا من خلالها مواردهم المشتركة؟

لم تتهرب دوفلو من ذلك السؤال الدقيق.[١٩] بل على العكس، فإنها وجدت تجربة طبيعية رائعة. ففي دولة ساحل العاج نجد الرجال والنساء في الأسرة يتشاركون المسئولية عن بعض المحاصيل. ولأسباب متعلقة بثقافة متجذرة يزرع الرجال والنساء محاصيل اقتصادية مختلفة (يزرع الرجال الكاكاو والبن ومحاصيل أخرى، وتزرع النساء موز الجنة وجوز الهند وأنواعًا قليلة من المحاصيل الأخرى). يكمن جمال هذه المنظومة من وجهة النظر البحثية في أن محاصيل الرجال ومحاصيل النساء تتجاوب مع أنماط هطول المطر بطرق مختلفة. ففي الأعوام التي تؤتي محاصيل الكاكو والبن أُكُلها جيدًا، يصير لدى الرجال المزيد من الدخل لينفقوه. وفي الأعوام التي تزدهر فيها محاصيل موز الجنة وجوز الهند، يتوافر لدى النساء المزيد من المال.

والآن نحتاج لأن نطرح سؤالًا دقيقًا، ألا وهو: هل يصير أطفال الأسرة في حال أفضل خلال الأعوام التي تزدهر فيها محاصيل الرجال، أم في الأعوام الأخرى التي تفيض فيها محاصيل النساء؟

الإجابة هي: عندما تؤتي محاصيل النساء أكلها، فإن النساء ينفقن بعضًا من فائض

الأموال لديهن على إطعام أسرهن. أما الرجال، فلا يفعلون ذلك. عذرًا يا رجال.

في عام ٢٠١٠ مُنحت دوفلو ميدالية جون بايتس كلارك. وهي جائزة تُمنح من قِبل الجمعية الاقتصادية الأمريكية لأفضل عالم اقتصاد دون سن الأربعين*. هذه الجائزة في نظر مهاويس الاقتصاد أرفع من جائزة نوبل في الاقتصاد؛ لأنها كانت تُمنح كل عامين (لكن ابتداءً من الدورة التي نالتها دوفلو عام ٢٠١٠ صارت تمنح سنويًّا). في أي حال كان تعتبر ميدالية كلارك هي نظير جائزة أفضل لاعب في الرابطة الوطنية لكرة السلة، لكنها تُمنح لأشخاص يرتدون نظارات سميكة (على سبيل التورية).

إن ما تفعله دوفلو هو تقييم لبرنامج. وصار عملها وعمل آخرين يستخدمون أساليبها يغيِّر حرفيًّا من حيوات الفقراء. ومن وجهة النظر الإحصائية، قد شجعتنا أعمال دوفلو على أن نوسع نطاق تفكيرنا عن كيف أن التجارب الضابطة – التي لطالما ظن فيها أنها من نطاق عمل العلوم المعملية – يمكن أن تُستخدم بشكل أوسع نطاقًا في إيجاد علاقات سببية في مَناحٍ عديدة أخرى من مناحي الحياة.

من له أن يعرف أي شيء عنك؟

في الصيف الماضي قمنا بتوظيف جليسة أطفال. وعندما قَدِمَت إلى المنزل، شرعتُ أقدِّم لها تعريفًا عامًّا بأسرتنا فقلت: «إنني أستاذ جامعي وزوجتي معلمة ...»

لوَّحَت جليسة الأطفال بيدها قائلة: «أعلم هذا، فقد تقصَّيت عنكم عبر محرك جوجل للبحث».

شعرت بالراحة لأنني لم أضطر إلى إكمال كلامي اللطيف ذلك، وبالانزعاج في الوقت نفسه من حقيقة أن حياتي كلها يمكن أن ترسم لها صورة كاملة من خلال بحث سريع على شبكة الإنترنت. إن قدرتنا على جمع كميات هائلة من البيانات وتحليلها – والتي هي تزاوُج بين المعلومات الرقمية مع القوة الحاسوبية الزهيدة الثمن وشبكة الإنترنت – هي قدرة متفردة في التاريخ البشري. إننا سوف نحتاج لقواعد جديدة لهذه الحقبة الجديدة.

الآن لنتأمل قوة البيانات من خلال مثال واحد من مؤسسة تارجت لتجارة التجزئة. وتارجت كأي شركة أخرى تكافح لزيادة أرباحها عن طريق فهْم زبائنها. ولإنجاز ذلك وظَّفت الشركة إحصائيين للقيام بنوع «التحليلات التنبؤية» التي وصفناها في محل سابق من الكتاب، وهم يستخدمون بيانات المبيعات مضافًا إليها معلومات أخرى عن المستهلكين، طلبًا لتحديدِ مَن يشتري ماذا ولماذا؟ لا نجد في هذا ما يضير، فهو يعني أنه إذا كان هناك متجر تارجت قرب منزلك، فمن المرجح أنك ستجد فيه ما تحتاج بالضبط.

* لم أكن مؤهلًا لنيل الجائزة عام ٢٠١٠ لأنني كنت قد جاوزت الأربعين من عمري، كما أنني لم أنجز شيئًا يجعلني أستحقها.

لكن لنتعمق قليلًا في مثال واحد عن نوع الأشياء التي يستطيع الإحصائيون، العاملون في قبوٍ لا نوافذ له في مقر الشركة الرئيسي، أن يتوصلوا إليها. توصلت تارجت إلى أن فترة الحمل هي فترة مهمة جدًّا فيما يتعلق بأنماط التسوق. فالنساء الحوامل يُقِمْنَ «علاقات بمتاجر التجزئة» قد تستمر لعقود. ونتيجة لهذا صارت تارجت تريد تحديد النساء الحوامل، وخاصة اللاتي في الثلث الثاني من حملهن، وأن تأتي بهن إلى متاجرها بشكل أكبر. قام أحد كتَّاب مجلة نيويورك تايمز بتتبع التحليلات التنبؤية لفريق تارجت بينما كانت تسعى لإيجاد النسوة الحوامل واجتذابهن.[٢٠]

كان الجزء الأول سهلًا جدًّا، فقد كانت تارجت تقوم بفتح قائمة تسجيل لهدايا حفلات المواليد الجدد، حيث تقوم الأمهات بكتابة الهدايا المطلوبة لأجل مولودها الجديد مقدمًا. أولئك النسوة هم أساسًا من متسوقي تارجت، وإنهن قد أبلغن المتجر بأنهن حوامل. لكن الحيلة الإحصائية تكمن في أن شركة تارجت قد توصلت إلى أن النساء اللاتي يتبعن نفس نمط تسوُّق زبائنها الحوامل هن أيضًا حوامل على الأرجح. فمثلًا عادة ما تتحول النساء الحوامل لاستخدام المستحضرات غير المعطرة، ويشرعن في شراء مكملات الفيتامينات، وأكياس كبيرة جدًّا من كرات القطن. قام عباقرة التحليلات التنبؤية في تارجت بتحديد خمسة وعشرين منتجًا تكوِّن معًا «سجل التنبؤ بالحمل». كان المقصد من ذلك التحليل هو أن تقوم المتاجر بإرسال بطاقات خصم للمنتجات التي تحتاجها الحوامل لأولئك السيدات؛ أملًا في أن يحولوهن إلى متسوقات دائمات لدى تارجت.

ما مدى كفاءة هذا النموذج؟ أوردت مجلة نيويورك تايمز ماجازين قصة عن رجل من منيابوليس دخل إلى متجر تارجت وطلب أن يقابل واحدًا من المديرين. كان الرجل يستشيط غضبًا لأن ابنته التلميذة في المدرسة الثانوية غمرتها بطاقات خصم لمنتجات خاصة بالحوامل من تارجت. سأل الرجل المدير قائلًا: «إنها ما زالت في المدرسة الثانوية، فترسلون لها بطاقات خصم على ملابس الرضع ومهودهم؟ هل تحاولون تشجيعها على أن تصير حبلى؟»

أخذ مدير المتجر يعتذر بحرارة للرجل، حتى إنه هاتفه بعد عدة أيام ليكرر اعتذاره. لكن في هذه المكالمة كان الرجل أقل غضبًا وجاء دوره في الاعتذار وقال لمدير المتجر: «لقد اتضح أنه قد جرى في منزلي ما لم أعي، سيحين موعد ولادتها في أغسطس».

لقد أدرك إحصائيو تارجت أن ابنته حامل قبل أن يدرك هو.

ذاك هو شأنهم ... وهو كذلك ليس من شأنهم. فقد يبدو ذلك تطفلًا يزيد عن اللزوم. ولهذا السبب صارت بعض الشركات تغطي على مقدار معرفتها بك. فمثلًا لو كنتِ امرأة في الثلث الثاني من حملكِ، فسوف تتلقين عبر البريد الإلكتروني بطاقات خصم على مهود الرضع والحفاضات، ومعها خصومات على دراجات جزِّ العشب وبطاقات جوارب لعبة

بولنج مجانية إذا ما اشتريتِ حذاءً للعب البولنج. قد تبدو لك تلك محض صدفة أن تأتيك بطاقات خصم لمنتجات الحوامل مع بقية المنتجات التي لا نفع منها. الحقيقة أن الشركة تعرف أنك لا تمارسين البولنج ولا تجزِّين حشائش حديقتك، وإنما تقوم بذلك بتغطية آثار معرفتها بكِ كي لا يبدو الأمر مثيرًا للجزع.

شركة فيسبوك، وهي الشركة التي ليس لها أية أصول مادية، قد صارت واحدة من أكثر الشركات قيمة على مستوى العالم. لكن بالنسبة إلى المستثمرين (على النقيض من المستخدمين) تمتلك فيسبوك أصلًا غاية في الضخامة، ألا وهو البيانات. لا يحب المستثمرون فيسبوك لأنه يجعلهم يعيدون التواصل مع من رافقوهم أو رافقوهن إلى حفل المدرسة الراقص، وإنما يحبونه لأن كل نقرة على فأرة الحاسوب تمنحهم بيانات عن المكان الذي يعيش فيه أولئك المستخدمون، وأين يتسوقون، وماذا يشترون، ومن يعرفون، وكيف يمضون أوقاتهم. أما بالنسبة إلى المستخدمين الآملين في التواصل مع من رافقوا أو رافقن إلى حفل المدرسة الراقص، فإن عملية جمع البيانات تلك يمكن أن تنتهك حدود الخصوصية.

ذكر كريس كوك نائب رئيس فيسبوك للإنتاج لصحيفة النيويورك تايمز: «إن التحدي الذي يبزغ في عصر المعلومات هو ما نفعله بتلك المعلومات»[٢١]

بالضبط.

وفي الفضاء العام يصير ذلك التزاوج بين البيانات والتكنولوجيا أكثر مخاتلة. فقد قامت الكثير من المدن حول العالم بتركيب الآلاف من كاميرات المراقبة الأمنية في الأماكن العامة، وبعض تلك الكاميرات سوف تمتلك عن قريب تكنولوجيا تسمح لها بتمييز الوجوه. تستطيع سلطات إنفاذ القانون أن تتبع أي سيارة كانت إلى أي مكان تذهب إليه (وتحتفظ بسجل مفصل لكل الأماكن التي ذهبت إليها)، عن طريق تركيب أداة تتبع عالمية فيها، ومراقبتها عبر الأقمار الصناعية. أتُعد هذه طريقة فعالة زهيدة الثمن لمراقبة أي نشاط إجرامي محتمل؟ أم إن الحكومة تستخدم التكنولوجيا كي تقمع حرياتنا الشخصية؟ قضت المحكمة الدستورية العليا بالولايات المتحدة عام ٢٠١٢، وبإجماع الآراء، أن الاحتمال الثاني هو الحادث، وحظرت على سلطات إنفاذ القانون تركيب أدوات التتبع تلك في السيارات الخاصة دون مذكرة قضائية*.

في نفس الوقت صارت حكومات كثيرة في جميع أنحاء العالم تحتفظ بقواعد بيانات ضخمة لبصمات الحمض النووي، والتي هي أدوات قوية في حل الجرائم. مَن الذي ينبغي أن توضع بصمة حمضه النووي في قاعدة البيانات تلك؟ أهي بصمات جميع المذنبين

---

* عرفت القضية باسم الولايات المتحدة ضد جونز.

المدانين؟ أهي بصمات كلِّ مَن يُلقى القبض عليه (سواء أُدِينَ بعدها أم لم يُدَنْ)؟ أو يجب أن تؤخذ عينة منَّا جميعًا؟

إننا ما زلنا في بداية نزالنا مع تلك القضايا التي تقبع عند التقاطع بين التكنولوجيا والبيانات الشخصية – وهي القضايا التي لم تكن محل نقاش أبدًا وقت أن كانت معلومات الحكومة تحفظ في خزائن موضوعة في أقبية ويعلوها الغبار لا في قواعد بيانات رقمية متاح للجميع أن يبحثوا عنها من أي مكان. لقد صار الإحصاء أهم عن ذي قبل؛ لأننا صرنا نحظى بفرص أكبر وأهم للاستفادة من البيانات. لكن رغم ذلك لن نعرف من المعادلات أيُّ استخدامات البيانات لائق وأيُّها ليس كذلك. فلا يمكن للرياضيات أن تحل محل حسن التقدير.

وفي هذا الإطار لنختم الكتاب ببعض التلازمات اللفظية: النار، السكاكين، السيارات، كريم إزالة الشعر. كل واحد من تلك الأشياء يخدم غرضًا مهمًّا. كلٌّ منها يجعل حياتنا أفضل. وكلٌّ منها قادر على أن يسبب مشكلات كبيرة إذا أسيء استخدامه.

والآن تستطيع أن تضيف الإحصاء إلى تلك القائمة. فلتقدم وتستخدم البيانات بحكمة، ولتستخدمها استخدامًا جيدًا.

ملحق

برامج إلكترونية إحصائية

أظنك لن تجري تحليلك الإحصائي مستخدِمًا قلم رصاص وورقة وآلة حاسبة. فيما يلي أصطحبك في جولة سريعة خلال حزمة البرامج الإلكترونية التي يشيع استخدامها لإنجاز تلك المهام التي تناولناها في هذا الكتاب.

مايكروسوفت إكسيل

قد يكون برنامج مايكروسوفت إكسيل أكثر البرامج استخدامًا عبر العالم في حساب إحصائيات بسيطة، كالمتوسط والانحراف المعياري. كما يمتلك إكسيل خاصية إجراء تحليل الانحدار الأساسي. غالبية أجهزة الكمبيوتر تباع محملة بحزمة برامج مايكروسوفت أوفيس؛ مما يعني أن مايكروسوفت إكسيل قد يكون متواجدًا على سطح مكتبك الآن. يعتبر إكسيل برنامجًا طيعًا في يد المستخدمين مقارنة بغيره من حزم البرامج الإلكترونية الإحصائية. ويمكن إجراء الحسابات الإحصائية الأساسية من خلال الخصائص الموجودة في شريط مهام البرنامج.

لكن إكسيل لا يمكنه أداء بعض المهام المتقدمة التي تقدر بعض البرامج التخصصية على القيام بها. لكن رغم هذا، هناك بعض الإضافات لإكسيل يمكنك أن تشتريها (أو تحملها مجانًا) سوف تزيد من قدرات البرنامج الإحصائية. من المزايا الكبيرة في إكسيل هو أنه يقدم سبلًا بسيطة لعرض البيانات الثنائية الأبعاد عن طريق رسوم بيانية خلابة من الناحية البصرية. وهذه الرسوم يمكن بكل سهولة أن تلصق في برنامجَي مايكروسوفت باوربوينت ومايكروسوفت وورد.

ستاتا*

هي حزمة برامج إحصائية يستخدمها الباحثون المتخصصون في جميع أنحاء العالم، وواجهة المستخدم فيها تتسم بمظهر أكاديمي رصين. لِستاتا نطاق واسع من الإمكانيات

* انظر: http://www.stata.com/.

التي تسمح لها بإنجاز مهام أساسية، مثل عمل جداول بيانات وحساب إحصائيات وصفية. ليس هذا بالطبع هو سبب استخدام أساتذة الجامعات والباحثين المتخصصين لِستاتا. صُممت هذه الحزمة الإحصائية للتعامل مع اختبارات إحصائية معقدة، ونماذج بيانات تتجاوز كثيرًا تلك الاختبارات والنماذج التي وصفناها في الكتاب.

تعتبر ستاتا خيارًا ممتازًا لأولئك الذين يتمتعون بفهم راسخ للإحصائيات (كما سيكون الفهم الأساسي للبرمجة مفيدًا في هذه الحالة)، وكذلك لأولئك الذين لا يتطلبون أمورًا مُبالغًا فيها – لا يريدون سوى أجوبة لاستفساراتهم الإحصائية. إذا ما كان هدفك هو الخروج برسوم بيانية سريعة من البيانات، فليست ستاتا هي مبتغاك. يقول المستخدمون الخبراء إن ستاتا قادرة على عمل رسوم بيانية لطيفة، لكن إكسيل قادر على تحقيق هذا الغرض بشكل أيسر.

تقدِّم ستاتا عددًا من حزمات البرامج الإلكترونية القائمة بذاتها. ولك أن تختار أنت أن تحصل على ترخيص لاستخدام البرنامج لمدة سنة واحدة (بعد انقضاء السنة لا يكون بإمكانك استخدامه على حاسوبك)، أو أن تحصل على ترخيص يستمر مدى الحياة. إحدى أزهد تلك الحزم ثمنًا هي Stata/IC، والتي صُممت لصالح «الطلبة والباحثين وزوِّدت بمجموعات بيانات معقولة الحجم». كما تقدِّم خصمًا لمن يعملون في الحقل التعليمي. لكن حتى حينها سيتكلف أي مستخدم يحتاج ترخيصًا لاستخدام Stata/IC لسنة واحدة ٢٩٥ دولارًا أمريكيًّا و٥٩٥ دولارًا أمريكيًّا للحصول على ترخيص دائم. إذا ما كنت تخطط لأن تطلق قمرًا صناعيًّا للمريخ، وكنت تحتاج لعمل حسابات معقدة، فعليك أن تلجأ إلى حزم برامج ستاتا المتقدمة، وهو ما سيكلفك آلاف الدولارات.

ساس*

يتمتع ساس بقبول واسع النطاق، ليس فقط من قِبل الباحثين المتخصصين، ولكن أيضًا من المحللين الماليين والمهندسين؛ وذلك بسبب اتساع نطاق إمكانياته. تتوافر من ساس حزمتان إحصائيتان مختلفتان. تسمى أولاهما ساس آناليتيكس برو، وهي القادرة على قراءة البيانات في أي شكل عملي والقيام بتحليل بيانات متقدم. كما يتسم هذا البرنامج بأدوات بصرية ممتازة، كإمكانيات رسم الخرائط المتقدمة. وهو ليس بالبرنامج الزهيد الثمن. حتى بالنسبة إلى من يعملون في حقل التعليم أو في القطاعات الحكومية، فالترخيص التجاري أو الشخصي الواحد يكلف ٨٥٠٠ دولار أمريكي، هذا بالإضافة إلى رسم الترخيص السنوي.

والحزمة الإحصائية الثانية هي ساس فيجوال داتا ديسكوفري. لهذه الحزمة

* انظر: http://www.sas.com/technologies/analytics/statistics/.

واجهة مستخدم سهلة الاستخدام، ولا تتطلب أي معرفة بالتشفير أو البرمجة، لكنها في الوقت نفسه تتيح إمكانيات متقدمة في تحليل البيانات. وكما يتضح من اسمها، فإن هذه الحزمة قد صُممت لتسمح للمستخدم بأن يستكشف البيانات عن طريق الوسائل البصرية التفاعلية. كما يمكنك أن تحول هذه البيانات المتحركة إلى عروض إلكترونية، أو صفحات على شبكة الإنترنت، أو غيرها من الوثائق. وهذه أيضًا ليست زهيدة الثمن. فأي ترخيص تجاري أو شخصي يكلف ٩٨١٠ دولارات أمريكية، بالإضافة إلى رسم الاستخدام السنوي.

كما تبيع ساس بعض الإمكانيات الإدارية المتخصصة، كمنتجات تستخدم الإحصائيات لاكتشاف التحايل والجرائم المالية.

آر

قد يبدو لك اسم هذه الحزمة البرامجية أشبه باسم شخصية من شخصيات أفلام جيمس بوند. لكن الواقع أن آر هي حزمة برامج إحصائية ذات قبول واسع، وهي مجانية أو «مصدر متاح». فهي قابلة لأن تُحمَّل وتفتح بكل سهولة على حاسوبك الشخصي في دقائق قليلة. كما أن هناك مجتمعًا نشطًا يسمى «مجتمع آر»، وفيه أشخاص يعرفون طرائقها ويستطيعون تقديم العون والمشورة لمن يحتاجها.

لا تُعَدُّ آر فقط الخيار الأقل تكلفة، وإنما تعتبر أيضًا واحدة من أكثر الحزم البرامجية التي ذكرناها مرونة. وهذه المرونة قد تراها أنت إما محبطة أو واحدة من أهم مزاياها. لو كنت محدثًا في مجال البرامج الإحصائية، فلن تقدم لك هذه الحزمة أي توطئة تقريبًا. ولن تساعدك واجهة المستخدم كثيرًا. لكن من جانب آخر سيجد المبرمجون (وحتى من يمتلكون معرفة أولية بمبادئ التشفير) انعدام التوطئة تلك تمنحهم حريتهم. فالمستخدمون يجدون كامل الحرية في أن يخبروا البرنامج ما يحتاجون له بالضبط، بما فيه إمكانية أن يجعلوا البرنامج يعمل مع برامج أخرى.

آي بي إم إس بي إس إس*

هذا برنامج يناسب الجميع، من الإحصائيين المخضرمين إلى المحللين الماليين حديثي العهد بالإحصاء. هو برنامج يناسب المبتدئين لأنه يقدم واجهة مستخدم بها قائمة سهلة الاستخدام. كما أنه يقدم عددًا من الأدوات أو «الوحدات» التي صُممت لأداء وظائف محددة، مثل آي بي إم إس بي إس إس فوركاستنج، وَآي بي إم إس بي إس إس أدفانسد ستاتستكس، وَآي بي إم إس بي إس إس فيجواليزاشن ديزاينر، وَآي بي إم إس بي إس إس رجرشن. يمكن شراء الوحدات مفردة أو مجتمعة في حزمة.

* انظر: http://www-01.ibm.com/software/analytics/spss/products/statistics.

لكن أكثر الحزم أساسية هي آي بي إم إس بي إس إس ستاتستكس ستاندرد إدشن، والتي تمكّنك من أن تحسب إحصائيات بسيطة وأن تجري تحليلًا أوليًّا للبيانات، كالتعرف على التوجهات الشائعة ونماذج التنبؤ. يكلف الترخيص التجاري الواحد الذي يستمر لفترة محددة ٢٢٥٠ دولارًا أمريكيًّا. أما الحزمة المتقدمة، والتي تحتوي على معظم الوحدات، فتكلف ٦٧٥٠ دولارًا أمريكيًّا. وهناك خصومات تقدَّم لمن يعملون في الحقل التعليمي.

المراجع

الفصل الأول: ما المقصد؟

1. Central Intelligence Agency, *The World Factbook*, https://www.cia.gov/library/publications/the-world-factbook/.
2. Steve Lohr, "For Today's Graduate, Just One Word: Statistics," *New York Times*, August 6, 2009.
3. Ibid.
4. Baseball-Reference.com, http://www.baseball-reference.com/players/m/mantlmi01.shtml.
5. Trip Gabriel, "Cheats Find an Adversary in Technology," *New York Times*, December 28, 2010.
6. Eyder Peralta, "Atlanta Man Wins Lottery for Second Time in Three Years," *NPR News* (blog), November 29, 2011.
7. Alan B. Krueger, *What Makes a Terrorist: Economics and the Roots of Terrorism* (Princeton: Princeton University Press, 2008).

الفصل الثاني: الإحصائيات الوصفية

1. U.S. Census Bureau, Current Population Survey, Annual Social and Economic Supplements, http://www.census.gov/hhes/www/income/data/historical/people/.
2. Malcolm Gladwell, "The Order of Things," *The New Yorker*, February 14, 2011.
3. CIA, *World Factbook*, and United Nations Development Program, *2011 Human Development Report*, http://hdr.undp.org/en/statistics/.
4. Baseball-Reference.com.

الفصل الثالث: وصف خادع

1. Robert Griffith, *The Politics of Fear: Joseph R. McCarthy and the Senate*, 2nd ed. (Amherst: University of Massachusetts Press, 1987), p. 49.
2. "Catching Up," *Economist*, August 23, 2003.
3. Carl Bialik, "When the Median Doesn't Mean What It Seems," *Wall Street Journal*, May 21-22, 2011.
4. Stephen Jay Gould, "The Median Isn't the Message," with a prefatory note and postscript by Steve Dunn, http://cancerguide.org/median_not_msg.html.
5. See http://www.movieweb.com/box-office/alltime.
6. Box Office Mojo (boxofficemojo.com), June 29, 2011.
7. Steve Patterson, "527% Tax Hike May Shock Some, But It's Only About $5," *Chicago Sun-Times*, December 5, 2005.
8. Rebecca Leung, "'The 'Texas Miracle': *60 Minutes II* Investigates Claims That Houston Schools Falsified Dropout Rates," CBSNews.com, August 25, 2004.
9. Marc Santora, "Cardiologists Say Rankings Sway Surgical Decisions," *New York Times*, January 11, 2005.
10. Interview with National Public Radio, August 20, 2006, http://www.npr.org/templates/story/story.php?storyId=5678463.
11. See http://www.usnews.com/education/articles/2010/08/17/frequently-askedquestions-college-rankings#4.
12. Gladwell, "Order of Things."
13. Interview with National Public Radio, February 22, 2007, http://www.npr.org/templates/story/story.php?storyId=7383744.

الفصل الرابع: الارتباط

1. College Board, FAQs, http://www.collegeboard.com/prod_downloads/about/news_info/cbsenior/yr2010/correlations-of-predictors-with-first-year-collegegrade-point-average.pdf.
2. College Board, 2011 College-Bound Seniors Total Group Profile Report,

http://professionals.collegeboard.com/profdownload/cbs2011_total_group_report.pdf.

3. See http://www.netflixprize.com/rules.

الفصل الخامس: الاحتمالات الأساسية

1. David A. Aaker, *Managing Brand Equity: Capitalizing on the Value of a Brand Name* (New York: Free Press, 1991).
2. Victor J. Tremblay and Carol Horton Tremblay, *The U.S. Brewing Industry: Data and Economic Analysis* (Cambridge: MIT Press, 2005).
3. Australian Transport Safety Bureau Discussion Paper, "Cross Modal Safety Comparisons," January 1, 2005.
4. Marcia Dunn, "1 in 21 Trillion Chance Satellite Will Hit You," *Chicago Sun-Times*, September 21, 2011.
5. Steven D. Levitt and Stephen J. Dubner, *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything* (New York: William Morrow Paperbacks, 2009).
6. Garrick Blalock, Vrinda Kadiyali, and Daniel Simon, "Driving Fatalities after 9/11: A Hidden Cost of Terrorism" (unpublished manuscript, December 5, 2005).
7. The general genetic testing information comes from Human Genome Project Information, DNA Forensics, http://www.ornl.gov/sci/techresources/Human_Genome/elsi/forensics.shtml.
8. Jason Felch and Maura Dolan, "FBI Resists Scrutiny of 'Matches,' " *Los Angeles Times*, July 20, 2008.
9. David Leonhardt, "In Football, 6 + 2 Often Equals 6," *New York Times*, January 16, 2000.
10. Roger Lowenstein, "The War on Insider Trading: Market Beaters Beware," *New York Times Magazine*, September 22, 2011.
11. Erica Goode, "Sending the Police before There's a Crime," *New York Times*, August 15, 2011.
12. The insurance risk data come from all of the following: "Teen Drivers," Insurance Information Institute, March 2012; "Texting Laws and

Collision Claim Frequencies," Insurance Institute for Highway Safety, September 2010; "Hot Wheels," National Insurance Crime Bureau, August 2, 2011.

13. Charles Duhigg, "What Does Your Credit Card Company Know about You?" *New York Times Magazine*, May 12, 2009.

الفصل الخامس والنصف: معضلة مونتي هول

1. John Tierney, "And behind Door No. 1, a Fatal Flaw," *New York Times*, April 8, 2008.
2. Leonard Mlodinow, *The Drunkard's Walk: How Randomness Rules Our Lives* (New York: Vintage Books, 2009).

الفصل السادس: مشكلات الاحتمالات

1. Joe Nocera, "Risk Mismanagement," *New York Times Magazine*, January 2, 2009.
2. Robert E. Hall, "The Long Slump," *American Economic Review* 101, no. 2 (April 2011): 431-69.
3. Alan Greenspan, Testimony before the House Committee on Government Oversight and Reform, October 23, 2008.
4. Hank Paulson, Speech at Dartmouth College, Hanover, NH, August 11, 2011.
5. "The Probability of Injustice," *Economist*, January 22, 2004.
6. Thomas Gilovich, Robert Vallone, and Amos Tversky, "The Hot Hand in Basketball: On the Misperception of Random Sequences," *Cognitive Psychology* 17, no. 3 (1985): 295-314.
7. Ulrike Malmendier and Geoffrey Tate, "Superstar CEOs," *Quarterly Journal of Economics* 124, no. 4 (November 2009): 1593-638.
8. "The Price of Equality," *Economist*, November 15, 2003.

الفصل السابع: أهمية البيانات

1. Benedict Carey, "Learning from the Spurned and Tipsy Fruit Fly," *New York Times*, March 15, 2012.

2. Cynthia Crossen, "Fiasco in 1936 Survey Brought 'Science' to Election Polling," *Wall Street Journal*, October 2, 2006.
3. Tara Parker-Pope, "Chances of Sexual Recovery Vary Widely after Prostate Cancer," *New York Times*, September 21, 2011.
4. Benedict Carey, "Researchers Find Bias in Drug Trial Reporting," *New York Times*, January 17, 2008.
5. Siddhartha Mukherjee, "Do Cellphones Cause Brain Cancer?" *New York Times*, April 17, 2011.
6. Gary Taubes, "Do We Really Know What Makes Us Healthy?" *New York Times*, September 16, 2007.

الفصل الثامن: مبرهنة النهاية المركزية

1. U.S. Census Bureau.

الفصل التاسع: الاستدلال

1. John Friedman, *Out of the Blue: A History of Lightning: Science, Superstition, and Amazing Stories of Survival* (New York: Delacorte Press, 2008).
2. "Low Marks All Round," *Economist*, July 14, 2011.
3. Trip Gabriel and Matt Richtel, "Inflating the Software Report Card," *New York Times*, October 9, 2011.
4. Jennifer Corbett Dooren, "Link in Autism, Brain Size," *Wall Street Journal*, May 3, 2011.
5. Heather Cody Hazlett et al., "Early Brain Overgrowth in Autism Associated with an Increase in Cortical Surface Area before Age 2 Years," *Archives of General Psychiatry* 68, no. 5 (May 2011): 467-76.
6. Benedict Carey, "Top Journal Plans to Publish a Paper on ESP, and Psychologists Sense Outrage," *New York Times*, January 6, 2011.

الفصل العاشر: استطلاعات الرأي

1. Jeff Zeleny and Megan Thee-Brenan, "New Poll Finds a Deep Distrust

of Government," *New York Times*, October 26, 2011.

2. Lydia Saad, "Americans Hold Firm to Support for Death Penalty," Gallup.com, November 17, 2008.
3. Phone interview with Frank Newport, November 30, 2011.
4. Stanley Presser, "Sex, Samples, and Response Errors," *Contemporary Sociology* 24, no. 4 (July 1995): 296-98.
5. The results were published in two different formats, one more academic than the other. Edward O. Lauman, *The Social Organization of Sexuality: Sexual Practices in the United States* (Chicago: University of Chicago Press, 1994); Robert T. Michael, John H. Gagnon, Edward O. Laumann, and Gina Kolata, *Sex in America: A Definitive Survey* (New York: Grand Central Publishing, 1995).
6. Kaye Wellings, book review in *British Medical Journal* 310, no. 6978 (February 25, 1995): 540.
7. John DeLamater, "The NORC Sex Survey," *Science* 270, no. 5235 (October 20, 1995): 501.
8. Presser, "Sex, Samples, and Response Errors."

الفصل الحادي عشر: تحليل الانحدار

1. Marianne Bertrand, Claudia Goldin, and Lawrence F. Katz, "Dynamics of the Gender Gap for Young Professionals in the Corporate and Financial Sectors," NBER Working Paper 14681, January 2009.
2. M. G. Marmot, Geoffrey Rose, M. Shipley, and P. J. S. Hamilton, "Employment Grade and Coronary Heart Disease in British Civil Servants," *Journal of Epidemiology and Community Health* 32, no. 4 (1978): 244-49.
3. Hans Bosma, Michael G. Marmot, Harry Hemingway, Amanda C. Nicholson, Eric Brunner, and Stephen A. Stansfeld, "Low Job Control and Risk of Coronary Heart Disease in Whitehall II (Prospective Cohort) Study," *British Medical Journal* 314, no. 7080 (February 22, 1997): 558-65.

4. Peter L. Schnall, Paul A. Landesbergis, and Dean Baker, "Job Strain and Cardiovascular Disease," *Annual Review of Public Health* 15 (1994): 381-411.
5. M. G. Marmot, H. Bosma, H. Hemingway, E. Brunner, and S. Stansfeld, "Contribution of Job Control and Other Risk Factors to Social Variations in Coronary Heart Disease Incidence," *Lancet* 350 (July 26, 1997): 235-39.

الفصل الثاني عشر: أخطاء الانحدار الشائعة

1. Gary Taubes, "Do We Really Know What Makes Us Healthy?" *New York Times Magazine*, September 16, 2007.
2. "Vive la Difference," *Economist*, October 20, 2001.
3. Taubes, "Do We Really Know?"
4. College Board, 2011 College-Bound Seniors Total Group Profile Report, http://professionals.collegeboard.com/profdownload/cbs2011_total_group_report.pdf.
5. Hans Bosma et al., "Low Job Control and Risk of Coronary Heart Disease in Whitehall II (Prospective Cohort) Study," *British Medical Journal* 314, no. 7080 (February 22, 1997): 564.
6. Taubes, "Do We Really Know?"
7. Gautam Naik, "Scientists' Elusive Goal: Reproducing Study Results," *Wall Street Journal*, December 2, 2011.
8. John P. A. Ioannidis, "Contradicted and Initially Stronger Effects in Highly Cited Clinical Research," *Journal of the American Medical Association* 294, no. 2 (July 13, 2005): 218-28.
9. "Scientific Accuracy and Statistics," *Economist*, September 1, 2005.

الفصل الثالث عشر: تقييم البرنامج

1. Gina Kolata, "Arthritis Surgery in Ailing Knees Is Cited as Sham," *New York Times*, July 11, 2002.
2. Benedict Carey, "Long-Awaited Medical Study Questions the Power of Prayer," *New York Times*, March 31, 2006.

3. Diane Whitmore Schanzenbach, "What Have Researchers Learned from Project STAR?" Harris School Working Paper, August 2006.
4. Gina Kolata, "A Surprising Secret to a Long Life: Stay in School," *New York Times*, January 3, 2007.
5. Adriana Lleras-Muney, "The Relationship between Education and Adult Mortality in the United States," *Review of Economic Studies* 72, no. 1 (2005): 189-221.
6. Kurt Badenhausen, "Top Colleges for Getting Rich," Forbes.com, July 30, 2008.
7. Stacy Berg Dale and Alan Krueger, "Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables," *Quarterly Journal of Economics* 117, no. 4 (November 2002): 1491-527.
8. Alan B. Krueger, "Children Smart Enough to Get into Elite Schools May Not Need to Bother," *New York Times*, April 27, 2000.
9. Randi Hjalmarsson, "Juvenile Jails: A Path to the Straight and Narrow or to Hardened Criminality?" *Journal of Law and Economics* 52, no. 4 (November 2009): 779-809.

الخاتمة

1. James Surowiecki, "A Billion Prices Now," *The New Yorker*, May 30, 2011.
2. Malcolm Gladwell, "Offensive Play," *The New Yorker*, October 19, 2009.
3. Ken Belson, "N.F.L. Roundup; Concussion Suits Joined," *New York Times*, February 1, 2012.
4. Shirley S. Wang, "Autism Diagnoses Up Sharply in U.S.," *Wall Street Journal*, March 30, 2012.
5. Catherine Rice, "Prevalence of Autism Spectrum Disorders," Autism and Developmental Disabilities Monitoring Network, Centers for Disease Control and Prevention, 2006, http://www.cdc.gov/mmwr/preview/mmwrhtml/ss5810a1.htm.

6. Alan Zarembo, "Autism Boom: An Epidemic of Disease or of Discovery?" latimes.com, December 11, 2011.

7. Michael Ganz, "The Lifetime Distribution of the Incremental Societal Costs of Autism," *Archives of Pediatrics & Adolescent Medicine* 161, no. 4 (April 2007): 343-49.

8. Gardiner Harris and Anahad O'Connor, "On Autism's Cause, It's Parents vs. Research," *New York Times*, June 25, 2005.

9. Julie Steenhuysen, "Study Turns Up 10 Autism Clusters in California," *Yahoo! News*, January 5, 2012.

10. Joachim Hallmayer et al., "Genetic Heritability and Shared Environmental Factors among Twin Pairs with Autism," *Archives of General Psychiatry* 68, no. 11 (November 2011): 1095-102.

11. Gardiner Harris and Anahad O'Connor, "On Autism's Cause, It's Parents vs. Research," *New York Times*, June 25, 2005.

12. Fernanda Santos and Robert Gebeloff, "Teacher Quality Widely Diffused, Ratings Indicate," *New York Times*, February 24, 2012.

13. Winnie Hu, "With Teacher Ratings Set to Be Released, Union Opens Campaign to Discredit Them," *New York Times*, February 23, 2012.

14. T. Schall and G. Smith, "Do Baseball Players Regress to the Mean?" *American Statistician* 54 (2000): 231-35.

15. Scott E. Carrell and James E. West, "Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors," National Bureau of Economic Research Working Paper 14081, June 2008.

16. Esther Duflo and Rema Hanna, "Monitoring Works: Getting Teachers to Come to School," National Bureau of Economic Research Working Paper 11880, December 2005.

17. Christopher Udry, "Esther Duflo: 2010 John Bates Clark Medalist," *Journal of Economic Perspectives* 25, no. 3 (Summer 2011): 197-216.

18. Esther Duflo, Michael Kremer, and Jonathan Robinson, "Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya," National Bureau of Economic Research Working Paper 15131, July 2009.

19. Esther Duflo and Christopher Udry, "Intrahousehold Resource Allocation in COte d'Ivoire: Social Norms, Separate Accounts and Consumption Choices," Working Paper, December 21, 2004.

20. Charles Duhigg, "How Companies Learn Your Secrets," *New York Times Magazine*, February 16, 2012.

21. Somini Sengupta and Evelyn M. Rusli, "Personal Data's Value? Facebook Set to Find Out," *New York Times*, February 1, 2012.

«كتاب بديع يجتذبك لقراءته».

– نيويورك تايمز

«غالبية من يهتمون بالرياضة والسياسة وإدارة الأعمال وغيرها من المجالات التي لا تُعَدُّ ولاتُحصَى والتي يتحكم فيها الإحصاء، سوف يستفيدون من هذا الكتاب الهام الجاذب الذي نجح في إيصال الفكرة التي استهدفها».

– فرانك نيوبورت، رئيس تحرير جالوب

«يجعل ويلان الإحصاء مشوقًا ممتعًا. ينزع كتابه عن الموضوع تعقيده، ويكشف ما تحته من مفاتن جذابة».

– مجلة ذا إيكونوميست

«ينزع ويلان عن الموضوع غطاءه الخارجي الذي لا لزوم له، ويكشف عن جماله الداخلي».

– هال فاريان، رئيس الباحثين الاقتصاديين بموقع جوجل.

تشارلز ويلان هو مؤلف «الاقتصاد عاريًا» وهو الكتاب الاكثر مبيعًا على مستوى العالم، والمراسل السابق لمجلة ذا إيكونوميست. كما أنه يدرِّس السياسة العامة والاقتصاد بمركز نلسون إيه روكفلر بكلية دارتموث، ويعيش في هانوفر بنيوهامبشاير مع أسرته.

٢٨٠ صفحة

بوك مانيا